Naval Research Laboratory

Washington, DC 20375-5000

AD-A242 822

Principles of Interpolator Design and Evaluation

ALAN SCHAUM

Advanced Concepts Branch
Optical Sciences Division

November 15, 1991

NOV 26 1991

91-16428

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE November 15, 1991	3. REPORT TYPE AND DATES COVERED Interim

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Principles of Interpolator Design and Evaluation	PE-62111N PR-RA11 W52 0W0

6. AUTHOR(S)
Alan P. Schaum

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Naval Research Laboratory Washington, DC 20375-5000	NRL Report 9356

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Office of Naval Technology Arlington, VA 22217-5000	

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution unlimited.	

13. ABSTRACT (Maximum 200 words)

It is shown that an error spectrum can be used to describe the performance of any convolutional interpolator operating on equally spaced points from an oversampled image. The error spectrum is linear in the image power spectrum and in an error factor that depends on only the interpolator and the distance from the sample points. The same form is shown to describe the interpolation of undersampled data, in an average sense. Simple forms are given for the error factor in either Fourier or real space, and standard interpolators are evaluated with them. Optimal interpolators are derived for various model image spectra: constant, Lorentzian, power law, and Gaussian. Practical methods of interpolator design are devised for use with image spectra that are known only partially or are not easily characterized analytically.

14. SUBJECT TERMS			15. NUMBER OF PAGES 45
Interpolation, Resampling, Frame differencing, Digital subtraction			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

CONTENTS

PRINCIPLES OF INTERPOLATOR
DESIGN AND EVALUATION

1. INTRODUCTION AND BACKGROUND

This report develops a general mathematical method for evaluating and optimizing any local interpolator that operates on equally spaced sample points. The performance metric is total squared error, for which a formula is derived that depends only on the underlying signal power spectrum, the interpolator tap weights, and the shift of the interpolation points. The method is used to evaluate standard interpolators, and it reveals previously unknown properties. For example: some interpolators perform better at large shifts than at small; also, the much maligned truncated SINC functions perform almost flawlessly at certain frequencies and are, furthermore, optimal for flat inband signal spectra.

The error formula also allows the design of interpolators that are optimal in a variety of contexts, examples of which are presented. These include perfect reproduction of selected frequencies; low-frequency optimum; hybrid combinations; and absolute error minimization for selected spectra.

An important application of interpolation is in the registration of image pairs, a process that is usually implemented in two steps. The first is the estimation of the relative shift or local distortion. Scene-based algorithms for this step have been developed [1] that can be accurate to better than one hundredth of a sample, depending on image statistics. Greater accuracy can be provided in some applications by direct measurement of the displacement. Residual registration error after using these displacement estimates is often dominated by inaccuracy in the second component of registration, the resampling of one of the images, which is interpolation at a discrete set of shifted points.

For remote sensing applications, registration "plays a crucial role in the correction of raw satellite image data" [2] collected by multiple sensors or at different times. The importance of algorithmic methods of image alignment then lies in the relaxation of optical/mechanical alignment tolerances [3], which is a vital cost/risk factor for satellite-based systems.

"In medical imaging, applications involving image registration are expanding rapidly: ... computed tomography, single-photon-emission computed tomography, positron-emission tomography, magnetic resonance imaging, nuclear medicine, ultrasound, and thermography" [4]. In digital subtraction angiography, images are compared before and after injection of X-ray absorbing elements, and patient or organ motion causes relative translation and rotation between successive images, which then require registration and resampling for comparison [5]. Improvements in the quality of such images can allow "reduction of injected contrast agents or X-ray doses" [6]. Here the term "resampling" means interpolation at points located at a fixed shift s from neighboring samples, rather than at the continuum of such points. This distinction is maintained in the analysis that begins in Section 2.

The detection of changes in a pair of images is also the basis for monitoring geologic, agricultural, and oceanic evolution in ecological and planetary studies. In military applications, autonomous surveillance and search and track systems often base the detection of moving targets on digital background subtraction followed by thresholding. The accuracy of all the above applications ultimately can hinge on the fidelity of the interpolator.

Interpolation is also required for general-purpose digital image processing, for motion generation, to scale images for display purposes, and to correct for slant aspects. It also has many traditional signal processing applications: speech processing, frequency multiplexing of single sideband systems, digital beamforming [7], time delay estimation, and data compression.

Progress in interpolator design can be traced to the former use of a suboptimal polynomial interpolator called Cubic Convolution [8] to reconstruct Landsat digital imagery [9]. This method was shown [10] to be a special instance of a class of four-point interpolators that was named Parametric Cubic Convolution (PCC), and the optimal value of the relevant parameter α was found to differ from that in popular use. The meaning of optimality depends on context. In Ref. 10, values of α were found that minimized a particular error measure:

(a) in an image-independent sense, and
(b) "absolutely," for an arbitrary but known power spectrum of the signal to be resampled.

For (a), image energy was assumed to be concentrated at low spatial frequencies. All members of PCC have perfect dc ($\nu = 0$) responses, and α was selected to enhance this low-frequency behavior by requiring the error measure to be a maximally flat function of frequency near dc. This low-frequency optimal choice of parameter value ($\alpha = -1/2$) supplanted the original choice ($\alpha = -1$), and eventually the name Cubic Convolution (CC) came to refer to PCC with the new choice of α. Here, we also use CC to mean PCC with $\alpha = -1/2$.

The error measure used in Ref. 10 is shown in Section 5 to be an average resampling error over all possible shifts of the resampling point. This is a reasonable measure for the *interpolation* problem, i.e., the construction of a dense set of resampled points. However, here we show how to find and minimize the error for any given value of the shift s, which is the distance of the resampled points from the nearest samples. We show further that the minimum-error four-point resampler in the sense of (a) above is indeed a cubic convolution, but it is not a member of PCC.

The function defined on the continuum by resampling an arbitrary signal at all shifts is called the interpolated function. The class of PCC interpolators was defined in part by constraining the interpolated function's values and first derivatives to be continuous. The imagery resulting from use of PCC to construct a continuum of resampled points has no steps or kinks. The interpolation function is smooth in the mathematical sense. Because of these constraints and the limited class of cubic interpolators allowed by even the full range of values of α, PCC is actually only a subclass of interpolators: Smooth Cubic Convolutions.

The optimal solution in the image-independent sense, above, for any fixed number N of sampling points is here called LF-N, for Low-Frequency optimal. The LF-N solutions generally can produce discontinuous first derivatives in interpolated imagery; when N is odd the functional values may be discontinuous as well. LF-N is smooth only in the limit $N \to \infty$.

When optimized in the absolute sense, PCC results in spectrum-dependent values of α, but the result is, of course, always a cubic interpolator. However, the absolutely optimal, i.e., minimum error, solution is not a polynomial type; this is described later.

2. STANDARD INTERPOLATORS

Here the philosophy of Parker et al. [5] and others is adopted, in which resampling is treated as a two-step process (Fig. 1):

- interpolation to form a function defined on the continuum, followed by
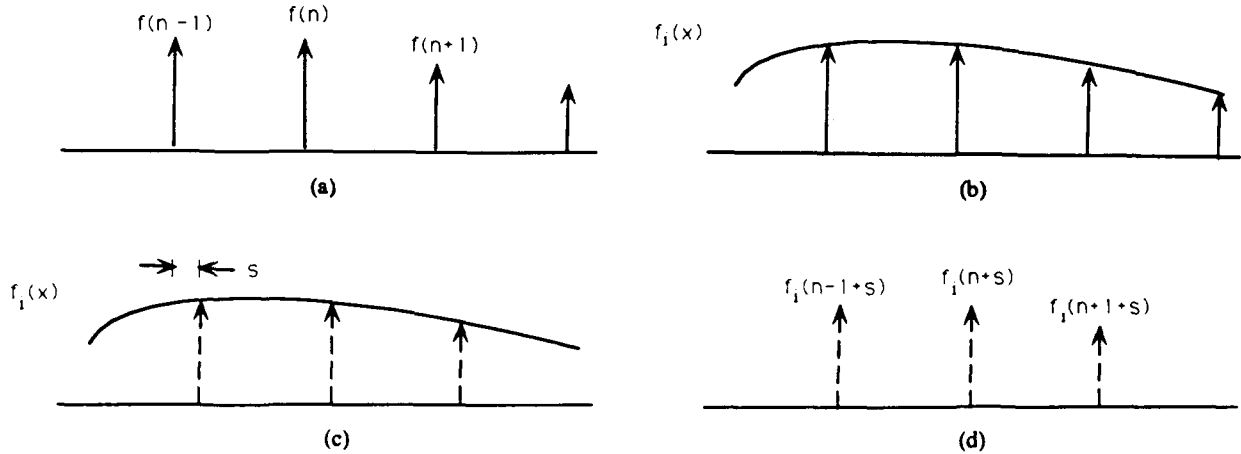- sampling the new function at whatever shift s is desired.

Fig. 1 — (a) Discrete set of sampled values from a continuous image; (b) interpolation, based on the sampled values, forming a continuous image; (c) sampling the interpolated image at a shift s to form: (d) a new set of resampled values

Letting $f(x)$ represent the underlying image intensity at position x, the sampled values of f, $f(n)$ (n an integer), are used to produce an interpolation estimate f_i of f:

$$f_i(x) = \sum_n r(x - n)f(n).$$
(1)

The kernel r defines the method of interpolation. The treatment here is in one dimension. Two-dimensional interpolation can be accomplished by a sequence of two one-dimensional operations.

Because the sampling grid is defined here to have unit spacing, the Nyquist frequency is $\nu_{Nyq} = 1/2$ cycles/sample. If the Fourier transform of f has no components with frequency $\nu \geq 1/2$, f is called "oversampled." According to the Nyquist Reconstruction Theorem, the kernel

$$r(x) = \text{SINC } (x) = \frac{\sin (\pi x)}{\pi x}$$
(2)

may be used in Eq. (1) to reproduce f exactly, that is $f_i = f$, whenever f is oversampled.

In practice, r must be of finite support (the range of x over which r is nonzero) to make Eq. (1) contain only a finite number of terms. For example, the SINC function may be truncated at $\pm N/2$. We call such an interpolator SINC-N. Two other common interpolators are nearest neighbor (NN) and linear (LIN). Figure 2 illustrates $r(x)$ for NN, LIN, and SINC. The supports for NN and LIN are 1 and 2, respectively. The SINC has infinite support, but when truncated at $\pm N/2$ to SINC-N, its support becomes N. Table 1 lists the corresponding analytic forms for r.
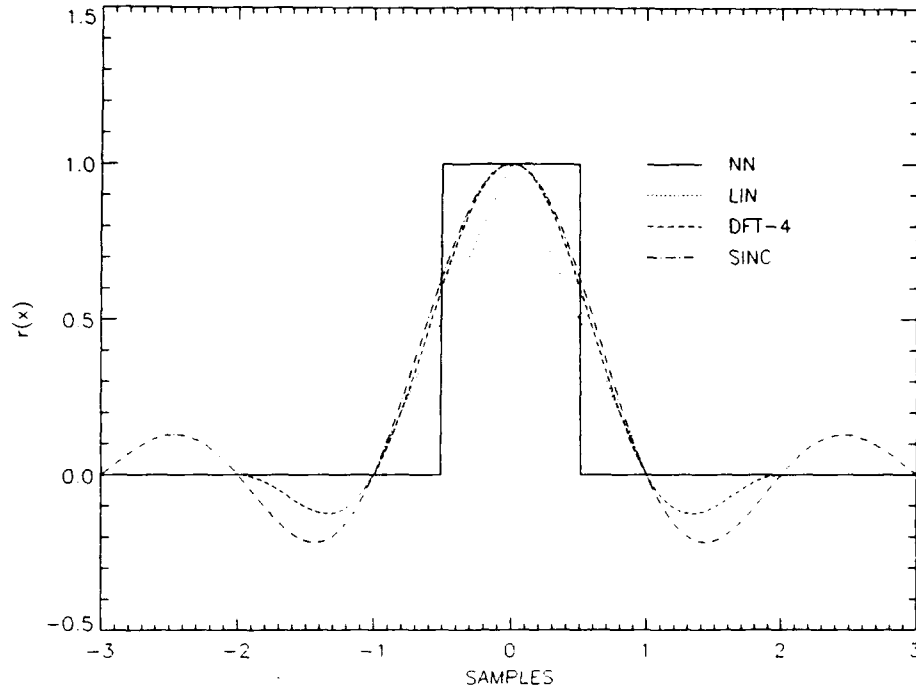
Fig. 2 — Interpolation kernels for Nearest Neighbor, Linear, DFT-4,
and SINC, with supports 1, 2, 4, and ∞, respectively

Table 1 — Common Interpolators (r, $\hat{r}$ are 0 where not specified)

Interpolator	$r(x)$	$\hat{r}(\nu)$
SINC ($\|x\| < \infty$)	$\dfrac{\sin(\pi x)}{\pi x}$	1 for $\|\nu\| < 1/2$
NN ($\|x\| < 1/2$)	1	$\text{sinc}(\nu)$
LIN ($\|x\| < 1$)	$1 - \|x\|$	$\text{sinc}^2(\nu)$
DFT-N ($\|x\| < N/2$) (N odd)	$\dfrac{\sin(\pi x)}{N \sin\left[\dfrac{\pi x}{N}\right]}$ *	$\displaystyle\sum_{k=-\frac{N-1}{2}}^{\frac{N-1}{2}} \text{sinc}(N\nu - k)$
(N even)	$\dfrac{\sin(\pi x)}{N \tan\left[\dfrac{\pi x}{N}\right]}$ *	$\displaystyle\sum_{k=-\left(\frac{N}{2}-1\right)}^{\frac{N}{2}-1} \text{sinc}(N\nu - k)$ $+\dfrac{1}{2}\left\{\text{sinc}\left[N\left(\nu - \dfrac{1}{2}\right)\right] + \text{sinc}\left[N\left(\nu + \dfrac{1}{2}\right)\right]\right\}$
PCC $\|x\| < 1$ $1 < \|x\| < 2$	$(\alpha+2)\|x\|^3 - (\alpha+3)\|x\|^2 + 1$ $\alpha(\|x\|^3 - 5\|x\|^2 + 8\|x\| - 4)$	$\dfrac{3}{(\pi\nu)^2}[\text{sinc}^2(\nu) - \text{sinc}(2\nu)]$ $+\dfrac{2\alpha}{(\pi\nu)^2}[3\,\text{sinc}^2(2\nu) - 2\,\text{sinc}(2\nu)] - \text{sinc}(4\nu)$

*These forms for DFT-N first appeared in Ref. 11.

Another interpolator whose performance has been studied with methods of experimental mathematics [12] is based on the discrete Fourier transform (DFT). The method is inspired by the Fourier-shift theorem. The idea is to compute the DFT, perhaps using modern fast Fourier transform algorithms, and then shift the phases of this finite, discrete transform as if it were the continuous Fourier transform. The result is then Fourier-inverted to produce a grid of resampled values. Figure 2 also illustrates DFT-4, which interpolates with four points. Table 1 lists the particularly simple analytic forms of the DFT kernels.

3. THE STANDARD DESCRIPTION OF INTERPOLATORS

The two most common ways of designing interpolators are to require similarity between:

- the kernel r and the SINC function, which is the perfect interpolator (of oversampled functions), and
- the Fourier transforms of r and SINC.

As an example of the first design method, the class PCC was defined not only by constraining r to be smooth like the SINC, but also by requiring $r(x)$ to agree with SINC(x) at all integral x. Also, the original cubic convolution interpolator corresponds to the parametric value $\alpha = -1$ in PCC; this choice equates the derivative of the kernel $r(x)$ to that of SINC at $x = \pm 1$. Other parametric choices can also be understood through similar interpretations in real space. Figure 3 compares PCC for several values of the parameter α to SINC, the ideal interpolator. The closeness of the $\alpha = -1$ curve to the SINC in the range $|x| \leq 1$ may explain the early preference for this parametric choice. Table 1 includes the analytic forms for PCC [10].

Analogous comparisons in Fourier space are often made between practical interpolators and the theoretical ideal. We use "^" to denote a Fourier transform, for example,

$$\hat{r}(\nu) = \int e^{-i\omega x} r(x)\,dx \quad (\omega = 2\pi\nu). \tag{3}$$

The transforms of the Fig. 2 interpolators are listed in Table 1 and plotted in Fig. 4 for $|\nu| \leq 3$. Among the examples shown, the transform of the SINC is the only one of truly finite frequency support, with unit value inband ($|\nu| < 1/2$) and zero in the sidebands ($|\nu| > 1/2$).

Interpolator design is often based on metrics associated with $\hat{r}$, such as inband cutoff, inband-to-sideband energy ratio, or weighted deviations (ripple) from the ideal SINC transform over some frequency range of interest. These approaches are rooted in an analogy with linear systems theory. Notice that the Fourier transform of Eq. (1) may be written as

$$\hat{f}_i(\nu) = \sum_n f(n)e^{-i\omega n}\hat{r}(\nu). \tag{4}$$

Therefore, $\hat{f}_i \propto \hat{r}$, which is the usual relationship between a linear filter and its output. However, the analogy is not complete because $\hat{f}_i$ is not also proportional to $\hat{f}$, the input to the filter. This can be traced to the sampling process, which introduces sideband structure and, in combination with interpolation, causes aliasing effects even if f is oversampled. Consequently, the relationship between standard frequency-domain fidelity measures of $\hat{r}$ and interpolator accuracy is often obscure. For example, if $f(x) = e^{i\omega_0 x + i\phi}$, then Eq. (4) becomes:

$$\hat{f}_i(\nu) = e^{i\phi}\sum_n e^{i(\omega_0 - \omega)n}\hat{r}(\nu) = e^{i\phi}\sum_n \delta(\nu - \nu_0 - n)\hat{r}(\nu), \tag{5}$$
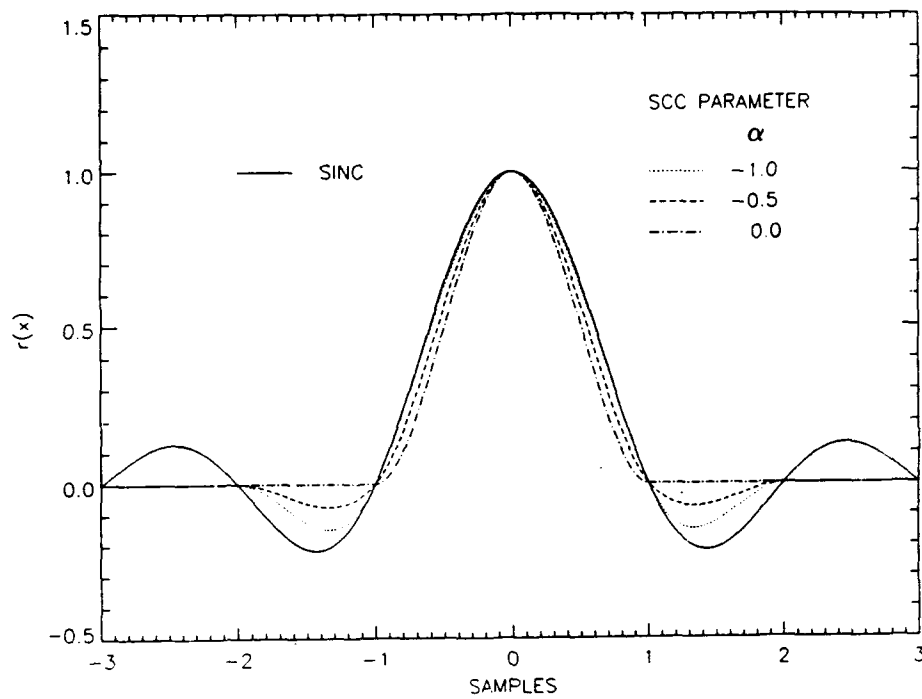
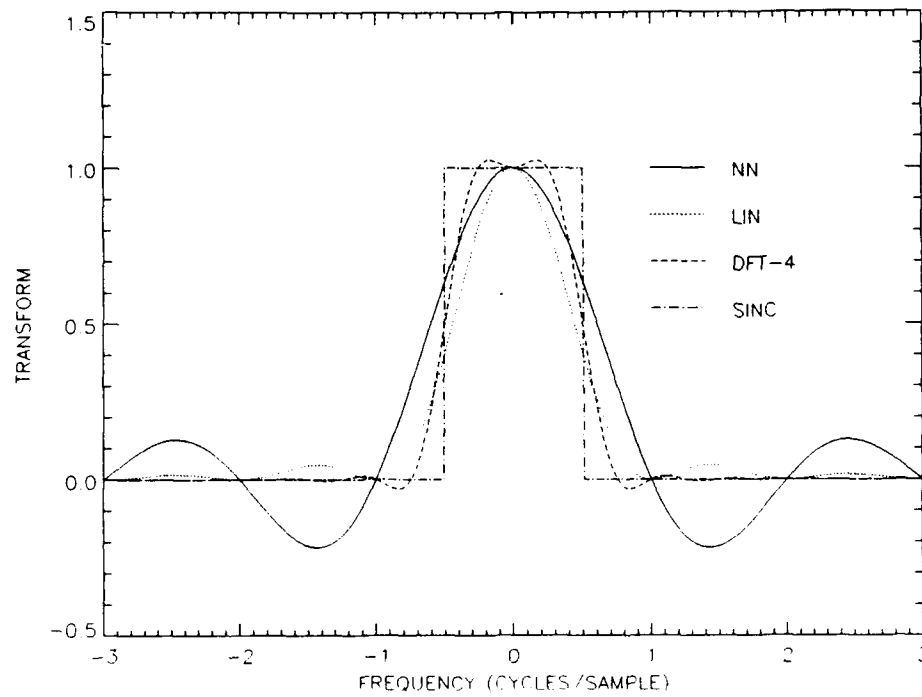Fig. 3 — Kernels for Parametric Cubic Convolution (PCC) compared to SINC

Fig. 4 — Transforms of the Fig. 2 kernels

with δ the Dirac delta function. (The last equality is justified in the next section.) Therefore, if r is to perfectly interpolate a sinusoid of frequency ν_0 $(= \omega_0/2\pi)$ (i.e., if $\hat{f}_i = \hat{f} = e^{i\phi}\delta(\nu - \nu_0))$, then Eq. (5) requires that

$$\delta(\nu - \nu_0) = \sum_n \delta(\nu - \nu_0 - n)\hat{r}(\nu) = \sum_n \delta(\nu - \nu_0 - n)\hat{r}(\nu_0 + n). \qquad (6)$$

From this, it follows that

$$\hat{r}(\nu_0 + n) = \delta_0^n, \qquad (7)$$

with δ_0^n the Kronecker delta. Therefore, interpolation with zero error at frequency ν_0 implies sideband as well as inband constraints on $\hat{r}$.

Furthermore, Eq. (7) implies that if r is designed to interpolate a *sideband* ν_0 perfectly, then $\hat{r}$ does *not* agree with the SINC transform at ν_0, which has the value zero. This reflects the failure of the Nyquist reconstruction formula in the sidebands. We show in Section 6 that kernels of finite support are often better than the SINC kernels in reproducing sideband signals.

All the interpolators of Fig. 4 satisfy Eq. (7) for $\nu_0 = 0$, meaning that a constant signal is interpolated perfectly. The transform of DFT-4, being the sum of regularly spaced SINC functions (see Table 1), satisfies Eq. (7) also for $\nu_0 = 1/4$. Generally, DFT-N perfectly interpolates $N/2$ frequencies spaced at intervals $1/N$ and starting at dc, provided that $\nu_0 = 0$ (dc) and $\nu_0 = 1/2$ (Nyquist) are each counted as "half" a frequency. For $\nu_0 \neq 0$, $1/2 + n$ (n is an integer), $\sin(\omega_0 x)$ and $\cos(\omega_0 x)$ can be reconstructed, in principle, from their sampled versions. But for $\nu_0 = 0$, only the cosine component can be present, and for $\nu_0 = 1/2 + n$, the sampling process cannot detect any sine component ($\sin(2\pi(1/2 + n)x) = 0$ for integer sample values x), which therefore cannot be reconstructed by any interpolator. Summarizing, DFT-4 perfectly interpolates any phase of sinusoids with frequencies $\nu_0 = 0$, $1/4$, as well as the cosine component of $\nu_0 = 1/2$.

As another example, DFT-5 interpolates $\nu_0 = 0$, $1/5$, and $2/5$ sinusoids perfectly. Section 7 shows that N-point interpolators can be constructed that perfectly reproduce an arbitrary linear combination of $N/2$ particular frequencies, with the above $\nu_0 = 0$, $(1/2 + n)$ counting convention. The DFTs are thus a particular subclass of these interpolators.

At frequencies that are not perfectly reproduced, the performance of DFT-4 or any other interpolator is not easily understood in terms of $\hat{r}(\nu)$. Note that this transform depends on $r(x)$ at all x, although the squared error at shifts s,

$$d_s^2 = \sum_n [f_i(n + s) - f(n + s)]^2, \qquad (8)$$

can depend on the interpolator kernel through only the specific tap weights $\{r(s + n)\}$. This is because the same is true for the sequence $\{f_i(n + s)\}$ (see Eq. (1)), which constitutes the resampled function.

Section 4 shows how to exploit this fact. A formula for d_s^2 is derived that depends explicitly on the kernel r through only the relevant tap weights, rather than through $\hat{r}(\nu)$. The formula provides a superior description of performance because in most applications only two to six tap weights are used, so that only a finite number of terms need to be computed to evaluate the error d_s^2. On the other hand, when expressed in frequency space, the error depends on an infinite number of values of the function $\hat{r}$ which is, furthermore, often difficult to compute analytically.

4. DERIVATION OF THE ERROR FORMULA

This section proves a fundamental theorem on the accuracy of any local interpolator. It also derives a formula for the error as a function of image power spectrum and kernel tap weights, $r(s + n)$. Several preliminary results are required, including a generalization of Parseval's Theorem.

A lemma central to the method used here is proved in the Appendix:

$$\text{comb}(x) \leftrightarrow \text{comb}(\nu). \tag{9}$$

The correspondence $\leftrightarrow$ relates a function to its Fourier transform. The comb function is defined by

$$\text{comb}(x) = \sum_n \delta(x - n), \tag{10a}$$

where δ is the Dirac delta function. Equation (9) says that the comb is its own Fourier transform. Thus, the transform of Eq. (10a)

$$\text{comb}(\nu) = \sum_n e^{-2\pi i n \nu} \tag{10b}$$

gives the relation used in Eq. (5).

Other correspondences for arbitrary functions g and h are: the Fourier shift theorem

$$g(x + s) \leftrightarrow e^{2\pi i s \nu} \hat{g}(\nu)$$

$$e^{-2\pi i x \nu_0} g(x) \leftrightarrow \hat{g}(\nu + \nu_0). \tag{11}$$

and the convolution theorem

$$g(x)h(x) \leftrightarrow (\hat{g} * \hat{h})(\nu)$$

$$(g * h)(x) \leftrightarrow \hat{g}(\nu)\hat{h}(\nu). \tag{12}$$

The definitions of convolution and of the comb (Eq. (10a)) may be used to prove

$$\sum_n g(x - n)h(n) = [g * (\text{comb} \cdot h)](x). \tag{13a}$$

A special case of Eq. (13a) is also useful:

$$\sum_n g(x - n) = (g * \text{comb})(x). \tag{13b}$$

These relations can be used to express a discrete sum of squared sample values of a function g in terms of its Fourier transform $\hat{g}$. Letting $g \to g^2$, $x \to 0$ in Eq. (13b), we can write

$$\sum_n g^2(n) = (g^2 * \text{comb})(0). \tag{14}$$

Letting $\mathscr{F}$ denote the Fourier transform operator, we successively apply Eqs. (12) and (9), the definition of $\mathscr{F}^{-1}$, and Eq. (12) again to simplify Eq. (14) further:

$$\sum_n g^2(n) = [\mathscr{F}^{-1}(\mathscr{F}(g^2 * \text{comb}))](0) = [\mathscr{F}^{-1}(\mathscr{F}(g^2) \text{ comb})](0)$$

$$= \int d\nu' [\mathscr{F}(g^2)(\nu') \text{ comb}(\nu')] = \int d\nu' [\hat{g} * \hat{g}](\nu') \text{ comb}(\nu'). \qquad (15)$$

Next, the frequency-space version of Eq. (10a) is used in Eq. (15), together with the definition of convolution:

$$\sum_n g^2(n) = \sum_n \int d\nu' \delta(\nu' - n) \int d\nu \hat{g}(\nu' - \nu)\hat{g}(\nu)$$

$$= \sum_n \int d\nu \hat{g}(n - \nu). \qquad (16)$$

If g is real, then

$$\hat{g}(\nu) = \hat{g}^*(-\nu) \quad \forall \nu, \qquad (17)$$

and so from Eq. (16),

$$\sum_n g^2(n) = \sum_n \int d\nu \hat{g}^*(\nu - n)\hat{g}(\nu). \qquad (18)$$

If, furthermore, the sequence $g(n)$ represents an oversampled version of g, so that $\hat{g}(\nu) = 0$ for $|\nu| \geq 1/2$, then Eq. (18) simplifies to

$$\sum_n g^2(n) = \int d\nu |\hat{g}(\nu)|^2, \qquad (19)$$

which is closely related to Parseval's Theorem. Note this distinction, however: in Eq. (19), $\hat{g}$ is the Fourier transform of the underlying continuous function g, not of the discrete Fourier series $\{g(n)\}$. These two transforms are equal if g is oversampled, but even if it is not, Eq. (19), which we will call the generalized Parseval's Theorem, still holds in an average sense described below.

Equation (19) implies the surprising fact that as long as g is oversampled, $\Sigma g^2(n)$ is independent of where the sampling grid is laid down relative to the image. Shifting the grid by an amount t in one direction is equivalent to shifting $g(x)$ in the other:

$$g(x) \rightarrow g(x + t). \qquad (20)$$

According to Eq. (11), in Fourier space Eq. (20) is equivalent to

$$\hat{g}(\nu) \rightarrow \hat{g}(\nu)e^{2\pi i t \nu}, \qquad (21)$$

under which change Eq. (19) is invariant. Because the value of Eq. (19) is independent of grid placement, we call it the strong form of the generalized Parseval's Theorem. When g is not oversampled, a weaker version of Eq. (19), described below, still holds.

Under the translation of the function g by the amount $t \in [0,1]$, according to Eq. (21), Eq. (18) becomes

$$\sum_n g^2(n) \rightarrow \sum_n e^{2\pi i t n} \int d\nu \, g^*(\nu - n) g(\nu). \tag{22}$$

Notice that the result of averaging Eq. (22) over t: $0 \rightarrow 1$ is Eq. (19). That is, the weak form of the generalized Parseval's theorem is also expressed by Eq. (19), if the left-hand side is understood to be averaged over all placements of the sampling grid.

In our application of Eq. (19), $\Sigma g^2(n)$ is d_r^2, which is a measure of interpolation error. The mean over t amounts to an average over grid translations; thus it is appropriate whenever the imagery of interest contains no preferred features relative to the sampling grid, which is the usual practical case.

To identify g with the interpolation problem, we let

$$g(x) = f_i(x + s) - f(x + s). \tag{23}$$

Then Eq. (8) allows the identification

$$d_s^2 = \sum_n g^2(n). \tag{24}$$

If the sampling grid were dense enough that $\{g(n)\}$ represented an oversampling of g, then Eq. (18) would become Eq. (19), so that

$$d_s^2 = \int d\nu \, |g(\nu)|^2, \tag{25}$$

in which, from Eqs. (23) and (11),

$$g(\nu) = [\hat{f}_i(\nu) - \hat{f}(\nu)] e^{2\pi i \nu s}. \tag{26}$$

Equation (25) also has strong and weak versions. If, for example, g is oversampled, then Eq. (25) is exact, without any averaging over grid translations. Unfortunately, for the interpolation problem, f_i, and hence g is almost never bandlimited. Therefore, g cannot be oversampled, even if f is. So the condition under which Eq. (25) has been shown to hold in the strong sense is not usually satisfied when g is given by Eq. (23).

Nevertheless, when Eq. (23) defines g, Eq. (25) is still exact whenever f is oversampled, even if g is not.

Using Eq. (13a) in Eq. (1) with $h = f$ and $g = r$, taking the transform, and applying Eq. (11) and Eq. (12) results in:

$$\hat{f}_i(\nu) = \hat{r}(\nu)[\text{comb} * \hat{f}](\nu), \tag{27}$$

which is usually not bandlimited because neither factor is, as shown below. Consequently, g in Eq. (26) is usually not bandlimited.

The frequency version of Eq. (13b) allows Eq. (27) to be rewritten as

$$\hat{f}_i(\nu) = \hat{r}(\nu) \sum_m \hat{f}(\nu - m). \tag{28}$$

The second factor is nonzero for arbitrarily large ν, as is the first whenever the kernel r is of finite support, i.e., for all practical interpolators. Consequently, f_i is not bandlimited, as claimed.

Nevertheless, using Eq. (18) together with Eqs. (26) and (28) results in

$$\sum_n g^2(n) = \sum_n \int d\nu [\hat{r}^*(\nu - n) \sum_m \hat{f}^*(\nu - n - m) - \hat{f}^*(\nu - n] \times \tag{29}$$

$$e^{2\pi i s m}[\hat{r}(\nu) \sum_j \hat{f}(\nu - j) - \hat{f}(\nu)],$$

which we now show to be independent of grid placement as long as f is oversampled.

Assuming then that

$$\hat{f}(\nu) = 0 \quad \text{for } |\nu| \geq \frac{1}{2}, \tag{30}$$

we examine the most complicated cross term of Eq. (29):

$$\sum_n e^{2\pi i s m} \int d\nu \hat{r}^*(\nu - n) \sum_m \hat{f}^*(\nu - n - m)\hat{r}(\nu) \sum_j \hat{f}(\nu - j). \tag{31}$$

Because of Eq. (30), the integration in Eq. (31) kills all summand terms except when $n + m = j$. So Eq. (31) becomes

$$\sum_n e^{2\pi i s m} \int d\nu \hat{r}^*(\nu - n) \sum_m \hat{f}^*(\nu - n - m)\hat{r}(\nu)\hat{f}(\nu - n - m). \tag{32}$$

Changing variables, $\nu \rightarrow \nu + n + m$, in Eq. (32) produces

$$\int d\nu |\hat{f}(\nu)|^2 \left\{ \sum_{n,m} e^{2\pi i s m}\hat{r}^*(\nu + m)\hat{r}(\nu + n + m) \right\}. \tag{33}$$

Each of the other cross terms of Eq. (29) produces a result similar to that of Eq. (33), but with a simpler expression within the braces. The final result

$$d_s^2 = \int |\hat{f}(\nu)|^2 e_s^2(\nu) d\nu, \tag{34}$$

with

$$e_s^2(\nu) = \sum_n e^{2\pi i n s} \left\{ \delta_0^n - \hat{r}^*(\nu - n) - \hat{r}(\nu + n) + \sum_m \hat{r}^*(\nu + m)\hat{r}(\nu + n + m) \right\} \tag{35}$$

is manifestly independent of grid placement. The quantity $|\hat{f}(\nu)|^2 e_s^2(\nu)$ is called the error spectrum. The error factor $e_s^2(\nu)$ depends only on the interpolating kernel, not on the image being interpolated. Equation (35) may be expressed more simply as

$$e_s^2(\nu) = |E_s(\nu)|^2, \tag{36}$$

with

$$E_s(\nu) = \sum_n [e^{2\pi i n s} \hat{f}(\nu + n)] - 1, \tag{37}$$

a "complex error" factor.

Summarizing, for any fixed shift s, the interpolation error is independent of an oversampled image's location relative to the sampling grid. Furthermore, the error is linear in the power spectrum. Power spectrum here simply means the squared modulus of the Fourier transform. It is not necessary to assume that $\{f(x)\}$ describes a stationary stochastic process, when the power spectrum equals the Fourier transform of the autocorrelation function. Equation (34) holds regardless of the stochastic character of f, including deterministic.

Notice that if s is a random variable that can be described only probabilistically, then the Fourier series in Eq. (35) can be used to express the moments of the error, which is also a random variable, because all the s dependence is contained in the first factor. For example, in the interpolation (as opposed to the resampling) problem, for which s may be considered uniformly distributed on $[0,1]$, the mean value of the error is just the $n = 0$ term of Eq. (35).

Also note that if f is undersampled, then Eq. (34) is still true in the average sense discussed earlier. The proof is similar to that used for the above weak form of the generalized Parseval's theorem. That is, the mean of d_s^2 over all grid placements depends only on the power spectrum of f and the error factor (Eq. (35)).

At this point, the dependence of the interpolation error on the image power spectrum has been isolated in Eq. (34), but the error factor remains expressed as an infinite sum over sidebands in Eqs. (36) and (37). The appearance of $\hat{f}(\nu)$ in Eq. (37) (and hence in Eq. (34)) at only the sideband frequencies $(\nu + n)$ is symptomatic of the complementary fact that d_s^2 can depend on the values of $r(x)$ only at the "side shifts" $(x = s + n)$. This latter fact means that as long as the values $r(s + n)$ are kept fixed, $r(x)$ for other x can be changed quite arbitrarily without changing the value of d_s^2, despite the attendant changes in $\hat{f}(\nu)$. This invariance is not obvious from Eq. (37) which is, moreover, impractical because realistic interpolators have infinite frequency-support. Consequently, the next logical step is to express $E_s(\nu)$ in terms of r instead of $\hat{r}$. We can expect to produce an expression containing only a finite number of terms whenever r is of finite support.

The first term of Eq. (37) may be written as

$$\sum_n e^{2\pi i n s} \hat{r}(\nu + n) = \sum_n e^{2\pi i n s} \int r(x) e^{-2\pi i x(\nu + n)} dx. \tag{38}$$

The real-space version of Eq. (10b) together with Eq. (10a) then permits Eq. (38) to be rewritten as

$$\int r(x) e^{-2\pi i x \nu} \text{comb}(s - x) dx = \sum_n e^{-2\pi i \nu(s+n)} r(s + n). \tag{39}$$

Therefore,

$$E_s(\nu) = \sum_n [e^{-i\omega(s+n)} r(s + n)] - 1, \tag{40}$$

which is the analog of Eq. (37) in terms of r instead of $\hat{r}$.

12

When expressed in real quantities only, Eqs. (36) and (40) become

$$e_s^2(\nu) = 1 - 2 \sum_n \cos (\omega[n + s])r(n + s) + \sum_m \cos (\omega m) \sum_n r(n + m + s)r(n + s), \quad (41)$$

which is the analog of Eq. (35).

Equations (40) and (36), together with the strong and weak interpretations of Eq. (34) permit the evaluation of any interpolator at shift s in terms of a finite number of kernel values whenever r has finite support. Should f have finite support, Eq. (37) plays a similar role to that of Eq. (40).

5. RELATION TO PRIOR WORK

The papers by Park and Schowengerdt [9,10] are seminal to this work. They considered an error measure

$$d^2 = \int_{-\infty}^{+\infty} [f_i(x) - f(x)]^2 dx. \quad (42)$$

Notice from Eq. (8) that

$$d^2 = <d_s^2>, \quad (43a)$$

where $< >$ denotes a mean value over an assumed uniformly distributed random shift s: $0 \rightarrow 1$. The error d^2 is thus appropriate when a (quasi) continuum of resampled values is of interest as, for example, in some image display applications.

Park and Schowengerdt [10] also derived an error factor $e^2(\nu)$ that can be interpreted similarly as the mean value of the error factor defined in this report. That is,

$$e^2(\nu) = <e_s^2(\nu)>. \quad (43b)$$

The mean value of the fundamental equation (Eq. (34)) is

$$d^2 = \int |\hat{f}(\nu)|^2 e^2(\nu) d\nu, \quad (44)$$

and the mean value of the error (Eq. (41)) can be written as

$$e^2(\nu) = 1 - 2\hat{r}(\nu) + \sum_m \cos (\omega m)(r * r)(m), \quad (45)$$

in which r is assumed to be symmetric (as in Ref. 10). Equation (45) can be called the fundamental formula of Ref. 10. It is accompanied by a fundamental theorem, namely the fundamental theorem that was derived in this report, but with the substitutions

$$e_s^2(\nu) \rightarrow e^2(\nu) \text{ and } d_s^2 \rightarrow d^2, \quad (46)$$

in Eq. (34).

This theorem was less surprising in Ref. 10 than it is here, because there it was shown to hold only in the above sense of an average over the shift s of the two sampling grids relative to each other. This should be distinguished from the grid placement averaging considered in the weak versions of the theorems discussed previously. That operation held the shift of one grid relative to the other fixed at the value s; the pair of grids was shifted as a unit relative to the image, or vice versa.

Also, in applications for which some performance criterion is defined in terms of d^2 or $e^2(\nu)$, Eq. (45) is normally useful only if an ansatz is first used for r (and, hence, also for $\hat{r}$). For example, d^2 can be minimized for a given power spectrum, or $e^2(\nu)$ can be set to zero at selected frequencies for some class of interpolators defined by the ansatz. In Ref. 10, the ansatz is PCC (Table 1). This allows the parametric calculations of $\hat{r}$ and $r * r$ which can then be inserted into Eq. (45), to be followed by the selection of that parameter α which minimizes d^2.

On the other hand, when Eq. (41) is used to constrain the error—at some fixed s—the resulting equations depend on the N kernel values $\{r(s + n)\}$, not on the transform of r or its self-convolution. Typically, this results in algebraic equations for these tap weights for any shift s, permitting the calculation of optimal kernels, rather than kernels that are optimal only in an averaged sense, and that are constrained by some ansatz. As noted earlier, in all optimality contexts considered in Ref. 10, the solutions for PCC are suboptimal because of the smoothness constraint defining the PCC ansatz. Notice from Table 1 that r for PCC is continuous, as is its first derivative, for all values of α.

6. PERFORMANCE EVALUATIONS

Examples

Figure 5(a) plots the error factors for some common interpolators at a .25 sample shift. The frequency 0.5 cycles/sample is the Nyquist frequency, above which no energy is present in an oversampled image. Generally at low frequencies, $e_s^2(\nu)$ varies as ν^2 for Nearest Neighbor, ν^4 for Linear, ν^6 for CC ("Cubic Convolution," i.e., PCC with $\alpha = -1/2$) and ν^2 for DFT-N interpolators. The small errors near dc are typical of polynomial methods. N-point interpolators that are designed optimally in this low-frequency region, called LF-N, have squared errors that vary as ν^{2N}. Nearest Neighbor interpolation is the same as LF-1, and LIN is LF-2. However, as noted earlier, CC is not LF-4, for which $e_s^2(\nu)$ is smaller at low frequencies, varying as ν^8.
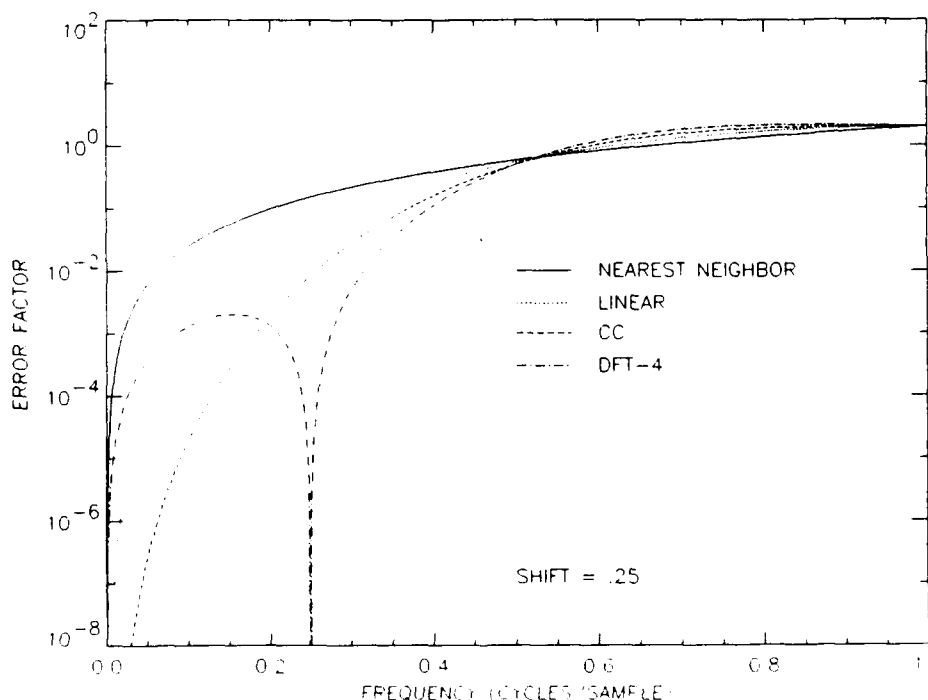
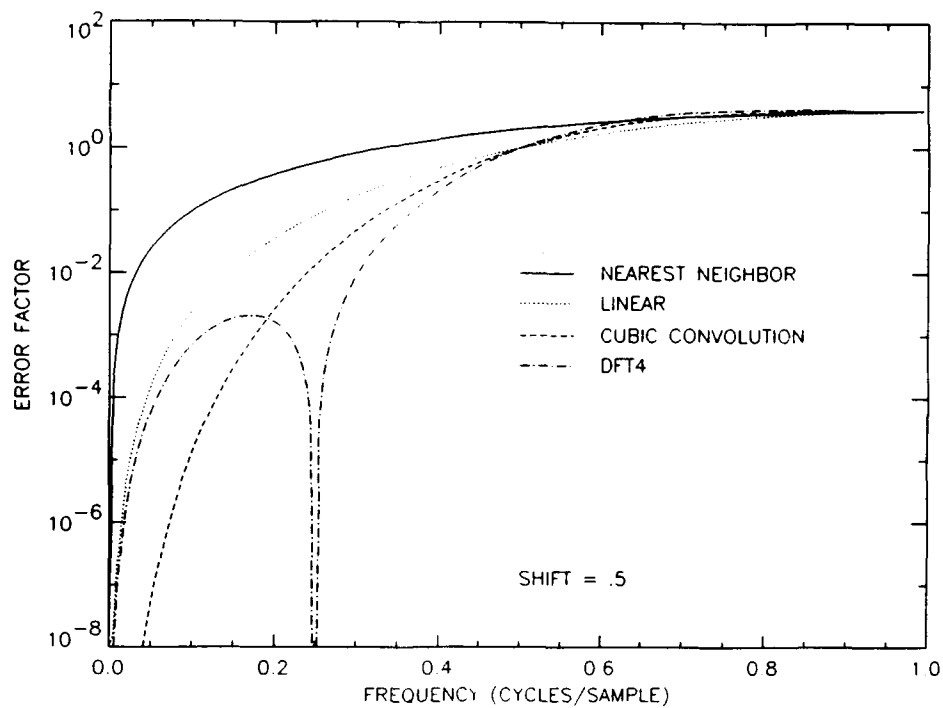Fig. 5(a) — Performance of standard interpolators (shift = .25 samples)

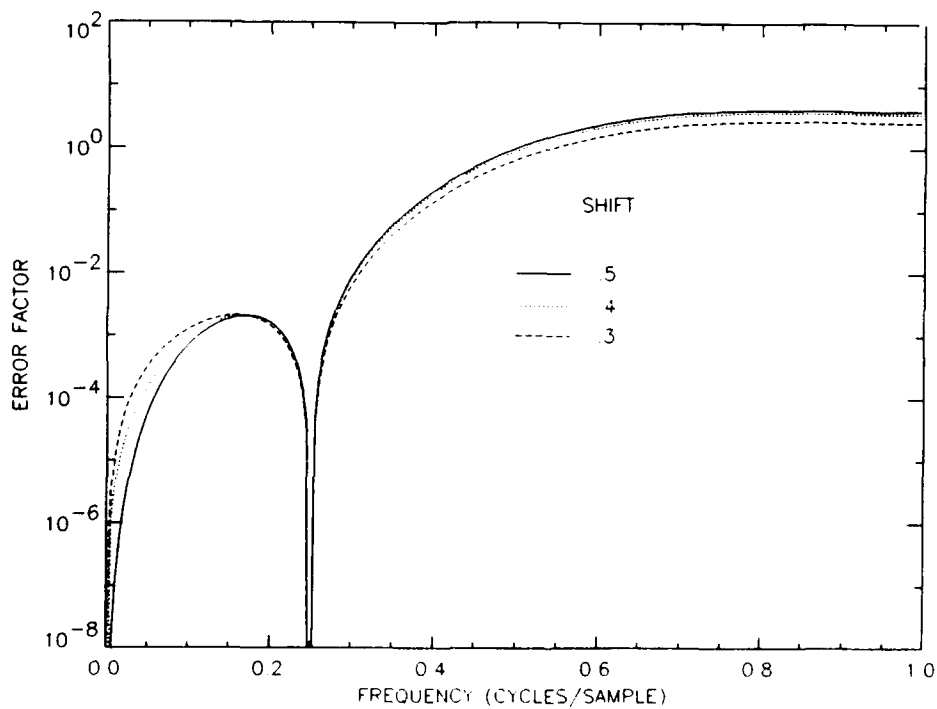Fig. 5(b) — Performance of standard interpolators (shift = .50 samples)

Fig. 5(c) — Performance of standard interpolators (DFT-4 at three shifts)

15

By contrast, a DFT interpolator of any order usually has a low-ν error that varies like ν^2, the same as for NN. An exception occurs at $s = 1/2$ (Fig. 5(b)) at which $e_{1/2}^2(\nu)$ improves for DFT at low frequencies, varying as ν^4, like LIN. Associated with this fact is the unusual property of DFT interpolators that they may perform better at the normally most stressing shift, $s = 1/2$, than at lower shifts. Figure 5(c) compares the performance of DFT-4 at three shifts. The plots show that, when applied to imagery with predominant energies below approximately .15 cycles/sample, DFT-4 errors actually increase as the shift decreases from .5 to .3 samples.

LF Interpolators

Energy in imagery is usually concentrated at low spatial frequencies, and the LF interpolators are designed to work well generally for image-like spectra. For LF-N, the error factor is constrained to be zero at dc, and the remaining degrees of design freedom, the number of which is determined by N, are used to make zero at $\nu = 0$ as many derivatives of the error as possible. This can be accomplished easily by requiring the complex error (Eq. (40)) to have zero derivatives. However, it is convenient first to multiply $E_s(\nu)$ by a phase factor $e^{i\omega s}$. Because of Eq. (36), this leaves unchanged the physically meaningful real error factor $e_s^2(\nu)$. Consequently, we can modify $E_s(\nu)$ of Eq. (40) to:

$$E_s(\nu) \rightarrow \sum_n e^{-i\omega n} r(s + n) - e^{i\omega s}.$$

(47)

The kth derivative of Eq. (47) at $\omega = 0$ is to be zero, meaning that

$$\sum_n (-n)^k r(s + n) = s^k.$$

(48)

It is clear from Eq. (1) that Eq. (48) requires r to be the interpolator that perfectly reproduces the polynomials s^k ($k = 0, ..., N - 1$) from a nearby set of N sample points. (We generally restrict s to the values $|s| \leq 1/2$ (N odd) and $0 \leq s \leq 1$, (N even).) This is the well-known Lagrange interpolator, which fits a unique ($N - 1$)-order polynomial to N points. Notice, however, that the Lagrangian polynomial so constructed is valid only on the central unit interval defined by the N sample points used to interpolate. As soon as the next unit interval is considered, a new sample point enters the finite sum in Eq. (48) and an old one leaves, so that the interpolated function f_i assumes the value of a new Lagrangian polynomial. This means that f_i, as well as its derivatives, can be discontinuous, a fact reflected in the Lagrangian kernels, $r(x)$. LF-1 is just NN, which has discontinuities at $x = \pm 1/2$. LF-2 is identical to LIN, which has derivative discontinuities at $x = 0$, ± 1. LF-3 has discontinuities at $x = 0$, $\pm 1/2$, and $\pm 3/2$, and LF-4 has derivative discontinuities at $x = 0$, ± 1, and ± 2.

Figure 6 illustrates these kernels for $N = 1$ through 4. The discontinuities at half-integers that appear in most interpolators with N odd usually make performance improvement marginal as $N \rightarrow N + 1$ with N even. The remainder of this report concentrates on comparisons with N even.

For positive arguments, the symmetric Lagrangian kernel is

$$r(n + s) = \frac{\prod_{\substack{j=n \\ \left[-\frac{N-1}{2}\right]}}^{\left[\frac{N-1}{2}\right]} (j+s)}{\prod_{\substack{j=n \\ \left[-\frac{N-1}{2}\right]}}^{\left[\frac{N-1}{2}\right]} (j - n)}, \quad 0 \leq n \leq \left[\frac{N-1}{2}\right], \quad \left[\frac{N}{2}\right] - \frac{N}{2} \leq s \leq \left[\frac{N}{2}\right] - \frac{N}{2} + 1 \quad (49)$$
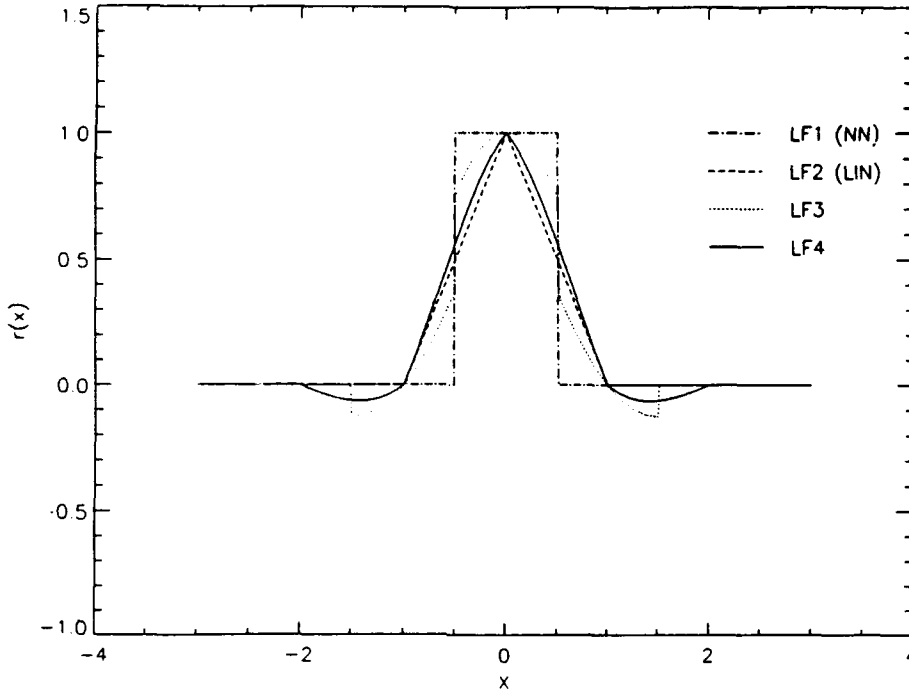
Fig. 6 — Kernels for Low-Frequency (LF) optimal interpolators

Here [ ] means "integer part of." Note that for integral q and $\epsilon \geq 0$, $[q + \epsilon] = q$, even if ϵq is negative.

Figure 7 compares the error factors (Eq. (41)) for LF interpolators to that of SINC at a standard shift $s = .25$. (Notice that the SINC has finite support in frequency space, making Eq. (37) rather than Eq. (40) more convenient for computing the plotted error.) Although $e_s^2(\nu)$ can be made flatter at $\nu = 0$ by using higher order interpolators, it becomes large beyond the Nyquist frequency. In the limit $N \to \infty$, LF-N approaches the "ideal" SINC interpolator. But clearly, the SINC is ideal only inband; for image energy beyond $\nu = 1/2$, the polynomial interpolators are superior.

7. OPTIMAL INTERPOLATORS

If the form of the image power spectrum is known, the results of Section 4 can be used to find the N-point interpolators that minimize the total squared error. For example, for a constant spectrum on $\nu \epsilon [-1/2, 1/2]$, the ν-integration from Eq. (34) can be calculated explicitly by using Eq. (41). The result is

$$d_2^2 = 1 - \sum_n r(n + s) \, \text{sinc} \, (n + s) + \sum_m \text{sinc} \, (m) \sum_n r(n + m + s) r(n + s)$$

$$(50)$$

$$= 1 - \sum_n r(n + s) \, \text{sinc} \, (n + s) + \sum_n r^2(n + m + s),$$

because sinc $(m) = \delta_0^m$. The sums in Eq. (50) are finite because $r(x)$ is zero for $|x| > N/2$.
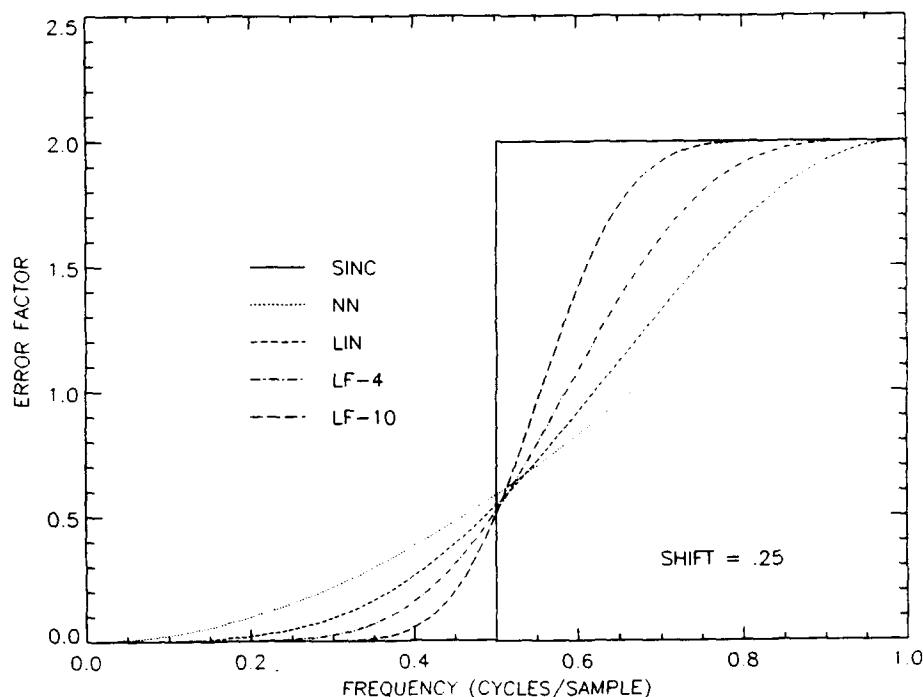
Fig. 7 — LF and SINC error factors compared

With the N tap weights $r(n + s)$ treated as independent variables, d_s^2 can be minimized by differentiation of Eq. (50). This leads to the solutions:

$$r(n + s) = \text{sinc } (n + s) \quad |n + s| \leq N/2. \tag{51}$$

Thus, SINC-N, the truncated sinc interpolator, which is much eschewed in the literature (Ref. 5, p. 35) because of the Gibbs' phenomenon in its Fourier transform, is in fact the optimal N-point interpolator for a flat inband image spectrum. Figure 8 plots the error factors for SINC-2,4,6. For a constant image spectrum, these can also be interpreted as the error spectra. Notice that the truncated SINCs reproduce certain frequencies almost perfectly. That is, the error plunges to near zero at certain frequencies for all shifts. This means that the continuous function f_i that results from interpolating from any set of sampled values of a sinusoid with one of these frequencies is nearly a copy of that sinusoid.

The oscillations in the interpolator transform that correspond to these dips in the error factor are often seen as undesirable. Figure 9 shows the result of applying a Hann window to SINC-6, which is a standard technique for reducing sidelobe oscillations. This windowing is just the multiplication of the kernel by a raised cosine to taper its edges and remove Gibbs' overshoot in the transform [13]. The figure shows the effect on the error factor. It also is smoother after windowing, which effects a trade in excellent reproduction of certain frequencies for a more uniform performance.
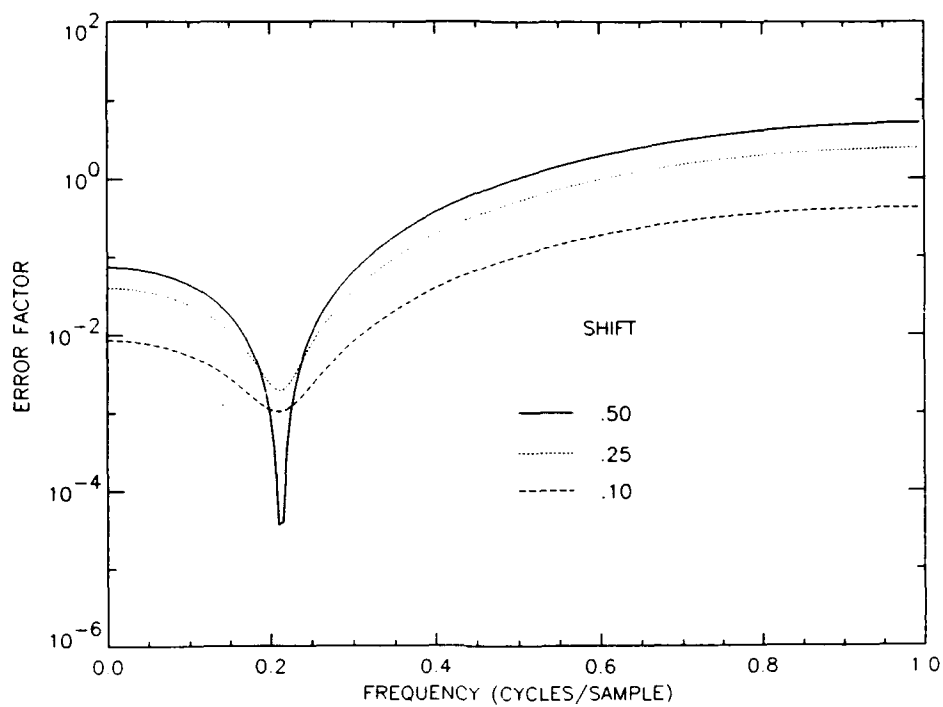
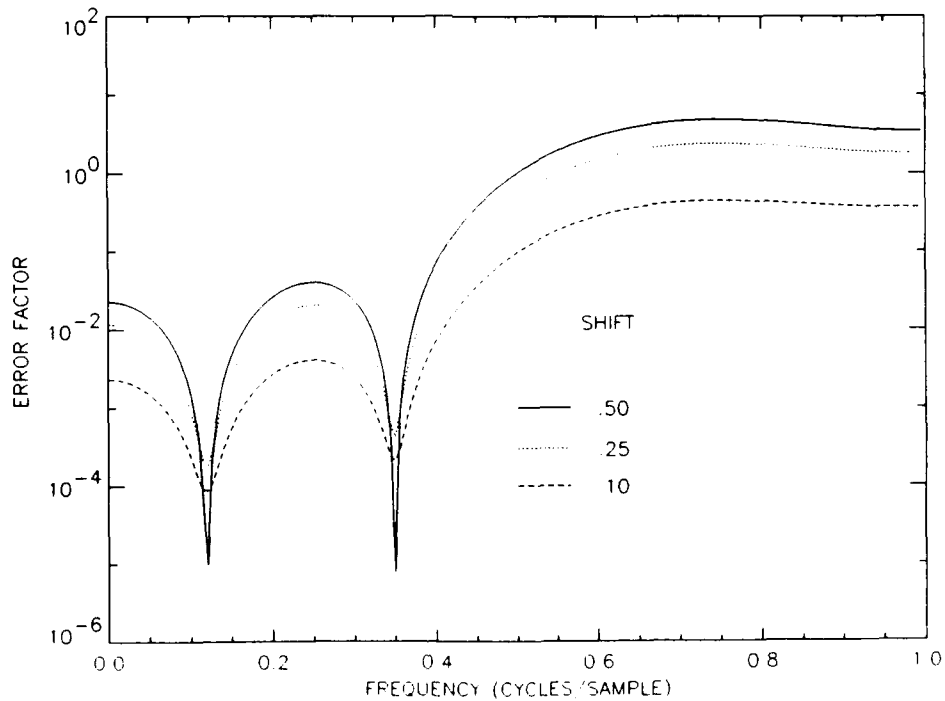Fig. 8(a) — Performance of truncated SINCs at three shifts: SINC-2

Fig. 8(b) — Performance of truncated SINCs at three shifts: SINC-4
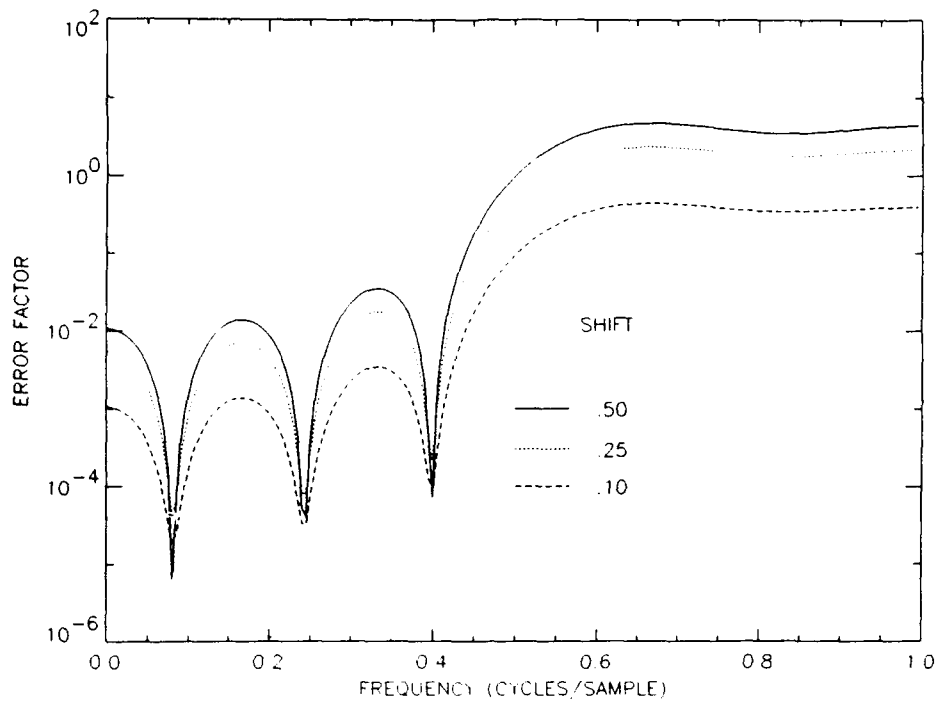
Fig. 8(c) — Performance of truncated SINCs at three shifts: SINC-6
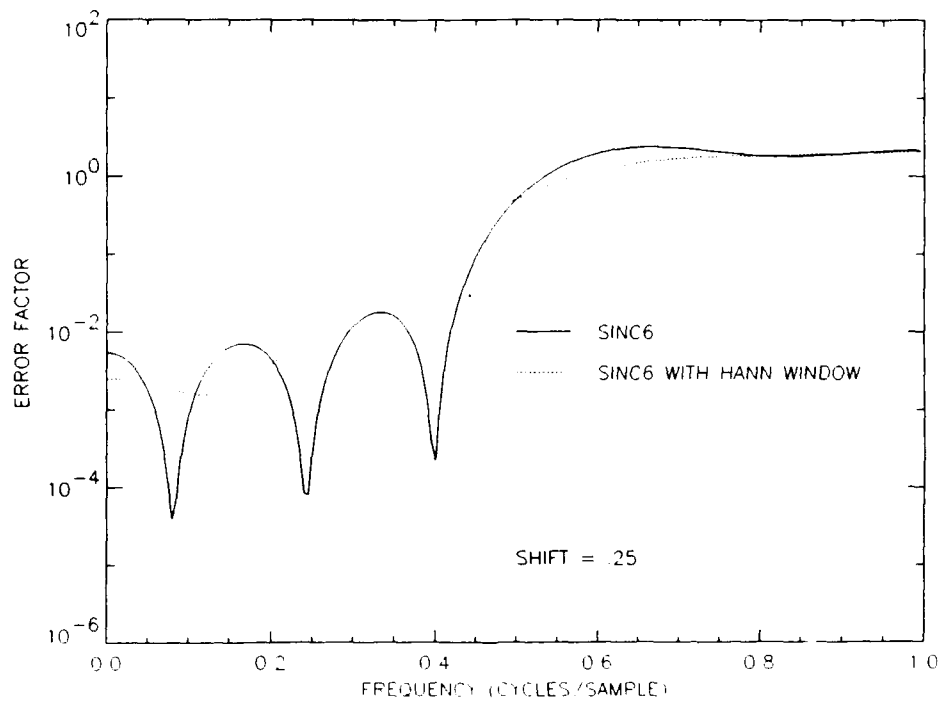

Fig. 9 — Effect of windowing SINC-6 with a Hann window

Figures 8 and 9 contain instances of a general problem with truncated kernels—imperfect dc response. This is often an undesirable feature that is remedied by renormalizing the kernel. For SINC-N, at any given shift s the replacement

$$\text{sinc } (n + s) \Rightarrow \frac{\text{sinc } (n + s)}{\sum_{m=\left[-\frac{N-1}{2}\right]}^{\left[\frac{N-1}{2}\right]} \text{sinc } (m + s)} \tag{52}$$

achieves the desired effect. Figure 10 shows the errors for $N = 6$ renormalized SINCs. They should be compared with Fig. 8(c). The lowest frequency local minimum has been shifted to dc; the second minimum has been shifted to the right; and the third's location is almost unchanged. The latter two minima are now also somewhat shift dependent, meaning that no longer is a continuous sinusoid at these frequencies especially well reproduced.
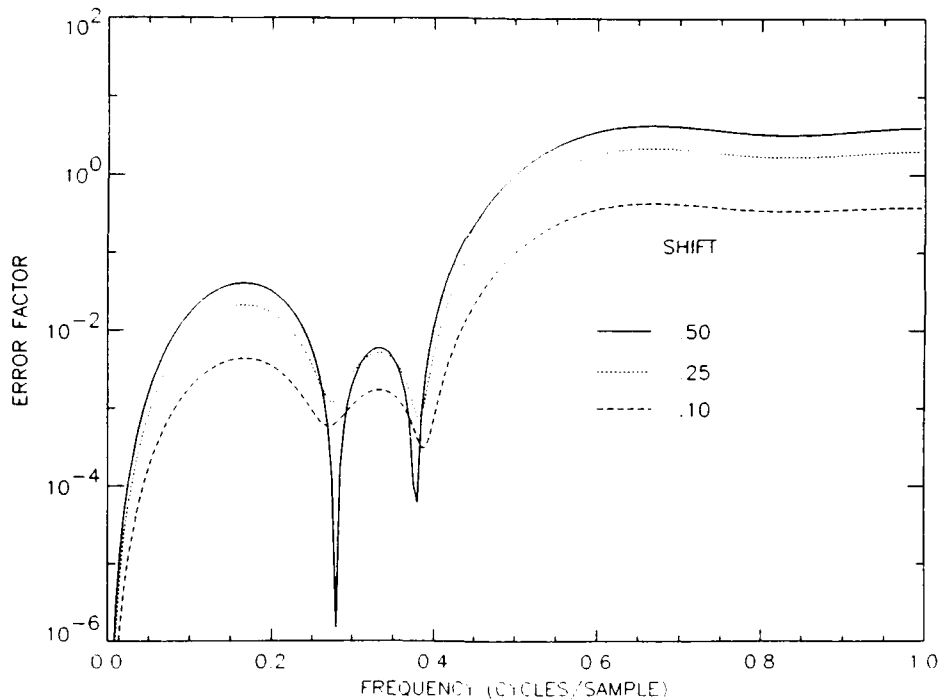

Fig. 10 — Effect of renormalizing SINC-6 to have zero dc error

Figures 8 through 10 show the ease with which the performance of common modifications of SINC-N, or of any other interpolator, can be predicted by using the formula in Eq. (41). Moreover, Eq. (34) has also been used to show that SINC-N is the minimum squared-error N-point interpolator for a flat Nyquist-sampled spectrum.

More generally, the minimum squared-error N-point interpolator can be derived, in principle, whenever $|\hat{f}(\nu)|^2$ is known up to an overall scale factor. Combining Eqs. (41) with (34) results in

$$d_s^2 = R(0) - 2 \sum_n R(n + s)r(n + s) + \sum_m R(m) \sum_n r(n + m + s)r(n + s), \qquad (53)$$

where

$$R(x) = \int_{-\infty}^{+\infty} \cos (x\omega)|\hat{f}(\nu)|^2 d\nu. \qquad (54)$$

Minimizing Eq. (53) with respect to the independent variables $r(n + s)$ results in an equation for the optimal N-point kernel

$$R(n + s) = \sum_m r(n + s - m)R(m), \qquad (55)$$

valid for those n for which $r(n + s)$ in Eq. (53) is nonzero, i.e., for the N values of n: $[-(N - 1)/2]$,...., $[(N - 1)/2]$.

A comparison of Eq. (55) with Eq. (1) shows that the former is equivalent to the requirement that r serve as a perfect interpolator for the function $R(x)$ for $x \in [-N/2, N/2]$, i.e., over the same support for which $r(x)$ is allowed to be nonzero. Equation (55) appeared in Ref. 2 as the condition minimizing a stochastic mean-squared-error. There the discrete sample values, here called $f(n)$, were assumed to represent an underlying stationary, ergodic, stochastic process. The function $R(x)$ was the autocorrelation function of that process, for which Eq. (54) is also valid if $|\hat{f}(\nu)|^2$ is interpreted as the power spectrum of a stochastic process with the above properties. Equation (55) has now been shown to hold even when the $f(n)$ result from the sampling of a deterministic function, if $R(x)$ is interpreted according to Eq. (54) instead of more restrictively, as the autocorrelation function of a random process.

Because $R(m)$ is symmetric, Eq. (55) represents a particular class of matrix equations called Toeplitz. General procedures [14,15] have been developed to solve these iteratively. Thus, in principle, $r(s + n)$ can always be found as a function of $2N$ numbers, $R(m)$ and $R(n + s)$, with $m = 0, ..., N - 1$ and $n = [-(N - 1)/2], ..., [-(N - 1)/2]$.

If these quantities are known, then Eq. (55) can be used to find the optimal $r(n + s)$. Notice that all the s-dependence in Eq. (55) is implicit, i.e., in the arguments of functions. This means that good interpolation may be possible without knowledge of s, as long as estimates of $R(m)$ and $R(n + s)$ are available. For example, if two images shifted by an unknown s are to be compared after one has been resampled, then R might be interpreted as an autocorrelation function, and periodograms [16] could be constructed from either image to estimate $R(m)$, and from both together to estimate $R(n + s)$. This technique introduces errors because of imprecision in the estimate of R. However, it completely eliminates errors from a prior step that we have ignored: estimation of the shift s. The competition between these errors depends on the performance of registration methods and is not studied here.

Lorentzian Spectrum

Because the optimization equations for the stochastic problem are identical in form to Eq. (55), prior results can be applied immediately to the present problem. One particularly surprising result occurs for a Lorentzian profile, that is a power spectrum of the form

$$|\hat{f}(\nu)|^2 \sim \frac{1}{\epsilon^2 + \nu^2} \qquad (\epsilon \geq 0). \qquad (56)$$

The transform of Eq. (56) is

$$R(x) \sim e^{-2\pi|x|\epsilon},$$ (57)

which can correspond to the autocorrelation function of a first-order autoregressive [AR(1)] stochastic process. Equation (56) is also a reasonable model for much natural imagery, independently of any probabilistic assumptions.

For the model in Eq. (56), the optimal N-point ($N \geq 2$) solution can be verified by substitution into Eq. (55) as:

$$r(s) = \frac{\rho^{s-1} - \rho^{1-s}}{\rho^{-1} - \rho}$$

$$r(s - 1) = \frac{\rho^{-s} - \rho^{s}}{\rho^{-1} - \rho}$$ (58)

otherwise, $r(s + n) = 0,$

with $\rho = e^{-2\pi\epsilon}$ and $0 \leq s \leq 1,$

which has support 2. That is, the optimal N-point solution is in fact a two-point interpolator! This curious result, first shown [2] for stochastic signals, also implies that for a $1/\nu^2$ power spectrum (the limit of Eq. (56) as $\epsilon \rightarrow 0$), the optimal N-point interpolator is just linear interpolation, which can be verified as the limiting form of Eq. (58).

Here we consider the general power-law spectrum

$$|\hat{f}(\nu)|^2 \sim \frac{1}{\nu^p}.$$ (59)

Figure 11(a) shows $e_s^2(\nu)|\hat{f}(\nu)|^2$, which is the power spectrum of the error, that is, of the difference between interpolated image and true value, for LIN, LF-4, and standard (smooth) Cubic Convolution (CC), for a $p = 2$ power spectrum, and at a shift of .25 samples. Although LIN is only a two-point interpolator, it has been claimed to be optimal for $p = 2$. That is, according to Eq. (34), the integral of the error spectrum is smaller for LIN than for any other interpolator, regardless of the value of N. Figure 11 shows how LIN accomplishes this feat. When considered only out to the Nyquist frequency ($\nu = 1/2$), LF-4 and CC are actually superior to LIN (by an rms factor of approximately 1.38). This remains true out to twice the Nyquist, over which frequency range CC and LF-4 are each a few percent better than LIN. It is only at higher frequencies (Fig. 11(b)) that LIN recovers from its inband inferiority. If Eq. (59) with $p = 2$ is assumed to hold for all frequencies, then LIN is a few percent better than either LF-4 or CC.

Figure 11(b) explains a counterintuitive result implied by the fact that LIN is optimal for $p = 2$. For those signals, including imagery, for which the $p = 2$ power spectrum is a reasonable model, experience shows that many polynomial interpolators outperform LIN. The reason this does not contradict the optimality claim for LIN is that the spectrum for which LIN produces the minimum error is usually an invalid model beyond the Nyquist frequency, when applied to imagery.
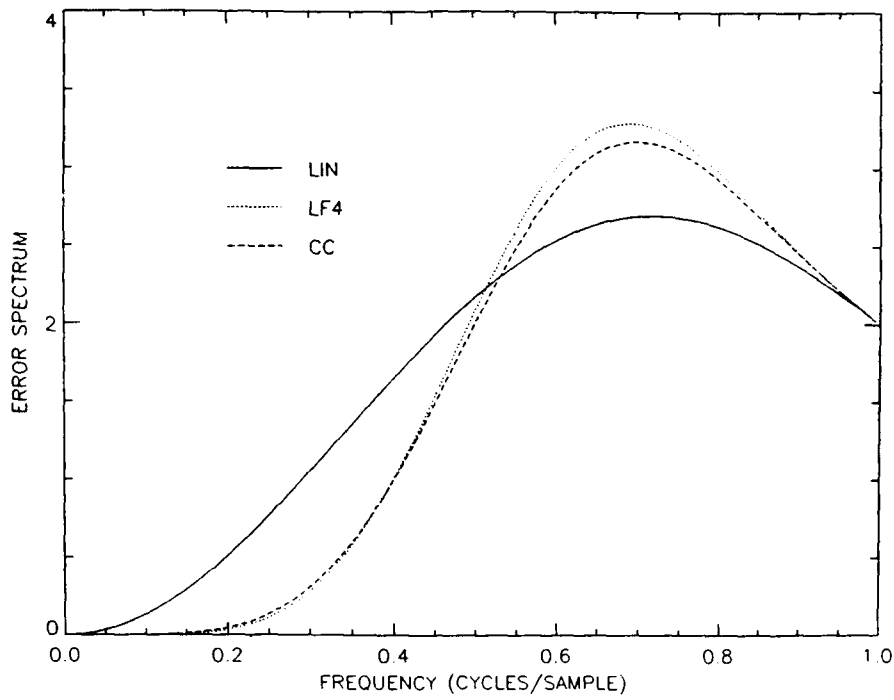
(a)
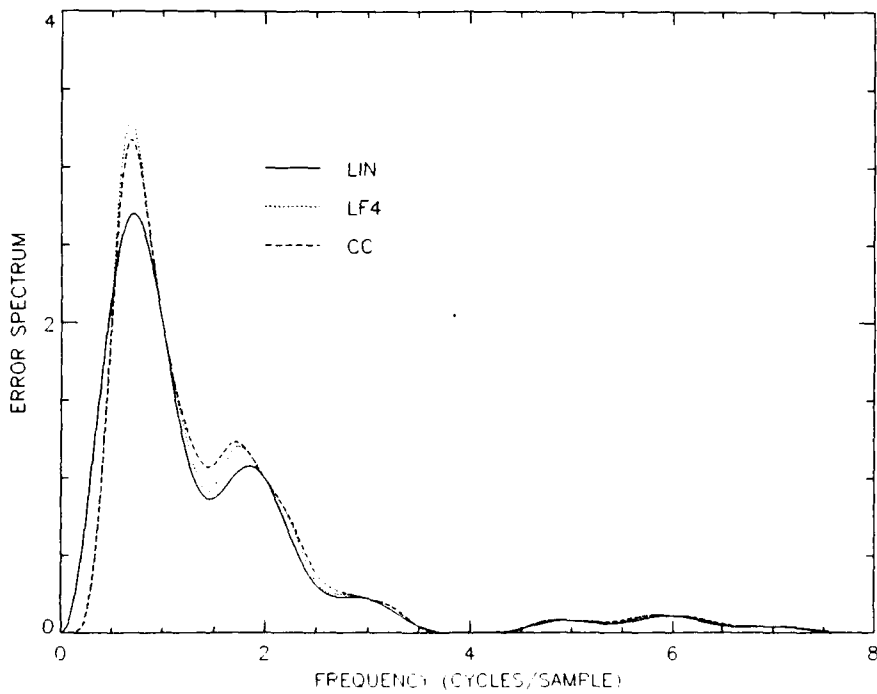
(b)

Fig. 11 — (a) Error spectra for LIN, LF-4, and cubic convolution (CC) for
$1/\nu^2$ image spectrum; (b) same as (a), but including higher frequencies

A power-law spectrum is commonly used to characterize one-dimensional slices of imagery, with p typically in the range $1 < p < 4$. Equation (59) is often assumed to hold in a piecewise sense, possibly with different values of p in different frequency ranges. Some low-frequency cutoff usually must be chosen to keep the total image energy (variance) finite, and commonly $\nu = 1/2$ (Nyquist) is chosen as a practical high-frequency cutoff. The p that is appropriate at low frequencies is often considered the most important because image energy tends to concentrate there. However, after interpolation, especially with polynomial kernels, the residual energy in a difference image often resides at higher frequencies, as exemplified in Fig. 11. Therefore, for extreme accuracy in interpolation, the high-frequency spectral content can be important also.

$p = 4$ Power Law Spectrum

As a second application of Eq. (55), we consider the case $p = 4$ in Eq. (59), which appears to be the limiting power of low-frequency divergence in natural imagery [17]. For $N = 2$, the solution to the Toeplitz equations is

$$r(s) = \frac{R(0)R(s) - R(1)R(1 - s)}{R^2(0) - R^2(1)}$$

$$r(s - 1) = \frac{R(0)R(1 - s) - R(1)R(s)}{R^2(0) - R^2(1)} \quad (0 \le s \le 1). \tag{60}$$

Although the integral in Eq. (54) defining R for the power spectrum of Eq. (59) diverges at low ν for $p = 4$, the power spectrum of the error $e_r^2(\nu)|\hat{f}(\nu)|^2$ is convergent at low frequency as long as

$$e_r^2(s) \sim \nu^{3+\delta}, \quad \text{with } \delta > 0. \tag{61}$$

This condition is violated by NN and (usually) DFT interpolators, for which $\delta = -1$; but even for LIN, the lowest nontrivial polynomial interpolator, $\delta = 1$. In practice, this means that the performance of NN and DFT can be highly dependent on the low-frequency cutoff that must exist in real imagery.

The explicit calculations of the kernel in Eq. (60) need a convergence mechanism for R, and so we solve instead for the power spectrum

$$|\hat{f}(\nu)|^2 \sim \frac{1}{\nu^4 + \epsilon^4}. \tag{62}$$

Then the limiting case $\epsilon \to 0$ will give the $p = 4$ power-law result. For Eq. (62), $R(x)$ in Eq. (54) can be evaluated in closed form:

$$R(x) \sim \exp\left[-\frac{|\epsilon x|}{\sqrt{2}}\right]\left[\cos\left[\frac{\epsilon x}{\sqrt{2}}\right] + \sin\left[\frac{|\epsilon x|}{\sqrt{2}}\right]\right]. \tag{63}$$

Then the minimum-error two-point interpolator for the power spectrum in Eq. (62) can be calculated from Eq. (60) by using Eq. (63). In the limit $\epsilon \to 0$, the solution (Eq. (60)) approaches linear, just as it does in the $p = 2$ case.

Similar calculations for $p = 4$ can be made for the four-point optimum. The Toeplitz solution for the power spectrum Eq. (62) is quite complicated, but the limiting solution for $\epsilon \rightarrow 0$ is simple:

$$r(s) = \frac{1}{5}(1 - s)(5 + 4s - 5s^2)$$

$$r(s + 1) = -\frac{1}{15}s(1 - s)(7 - 5s) \tag{64}$$

$$r(x) = 0 \quad |x| > 2$$

$$r(x) = r(-x) \quad \forall x.$$

Figure 12(a) shows the error factors at $s = .25$ for three four-point interpolators, and Fig. 12(b) shows the corresponding error spectra for a $p = 4$ input spectrum. Beyond $\nu = 1$, the curves are nearly identical, unlike the corresponding $p = 2$ curves. Nevertheless, it is again evident that the best interpolator (as measured by $\sqrt{d_s^2}$) depends on the frequency ν_t at which the $p = 4$ spectrum might be truncated to match realistic image data.

Integration of the Fig. 12(b) curves shows that over all frequencies ($\nu_t = +\infty$), the optimal solution (Eq. (64)) is only about 2 percent better than either LF-4 or CC; for $\nu_t = 1/2$, it is 8 percent and 9 percent better, respectively. The optimal is still superior for $\nu_t = 1/4$, but for $\nu_t = .1$, LF-4 produces 5.5 times less rms error than the ($\nu_t = +\infty$) optimal. Evidently from Fig. 12(b), $e_s^2(\nu)$ for the optimal varies as ν^4 for small ν, so that the error spectrum approaches a constant; for LF-4, $e_s^2(\nu)$ varies as ν^8.

Gaussian Spectrum

One final class of power spectra is considered. A single point-like object or an assortment of such that are spatially separated by several times the blur imposed by a sensor's optics corresponds to a power spectrum proportional to the square of the sensor transfer function. Often a good model of this function is a Gaussian, and so our final example is

$$|\hat{f}(\nu)|^2 \sim e^{-\omega^2\sigma^2}. \tag{65}$$

This corresponds to a system point spread function of

$$p(x) \sim e^{-(x^2/2\sigma^2)}. \tag{66}$$

We consider three cases: $\sigma = 1$, $1/2$, and $1/3$, measured in samples. The first case represents an imaging system that samples at nearly the Nyquist rate; the third represents one that undersamples. For example, a line of square detectors (Fig. 13(a)) used in a scanning imager with "matched optics"* corresponds approximately to $\sigma = 1/3$ sample in the cross-scan direction, and to $\sigma = 1$ sample in the in-scan if the continuous detector output is sampled at a rate of approximately three times per scan across a detector. In this situation, a cross-scan sample corresponds to one detector width; an in-scan sample corresponds to 1/3 of a detector width. Because the detectors are butted in the Fig. 13(a) example, $\sigma = 1/3$ also corresponds to the effective sampling rate in either direction along a mosaic array of closely spaced detectors (Fig. 13(b)) that image in a staring mode.

---

*Optics in which the first zero of a centered point response falls at the detector edge.
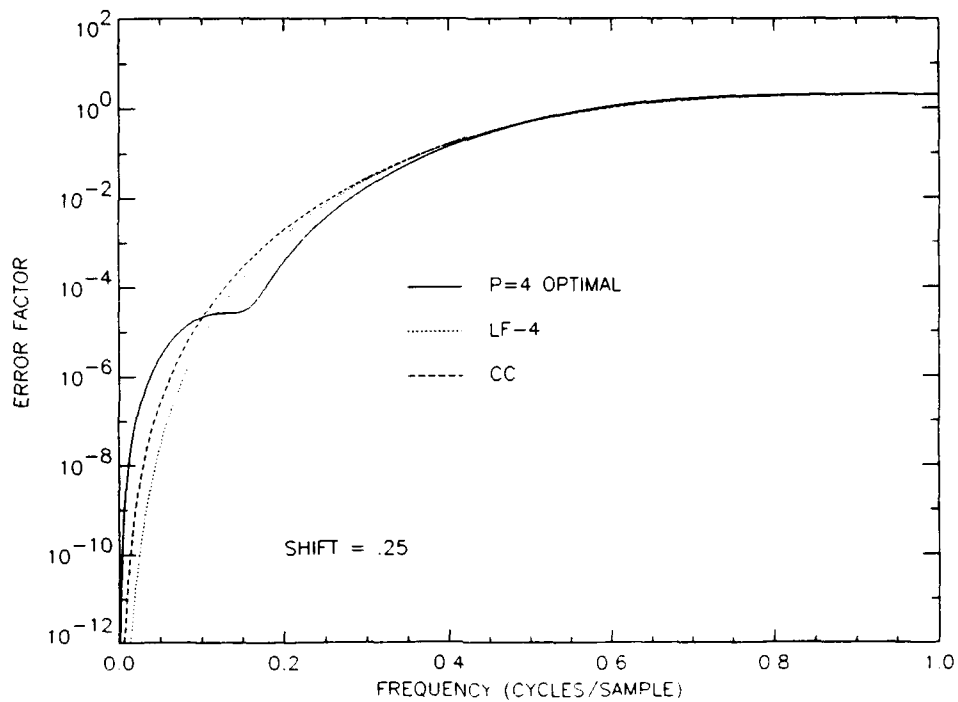
(a)
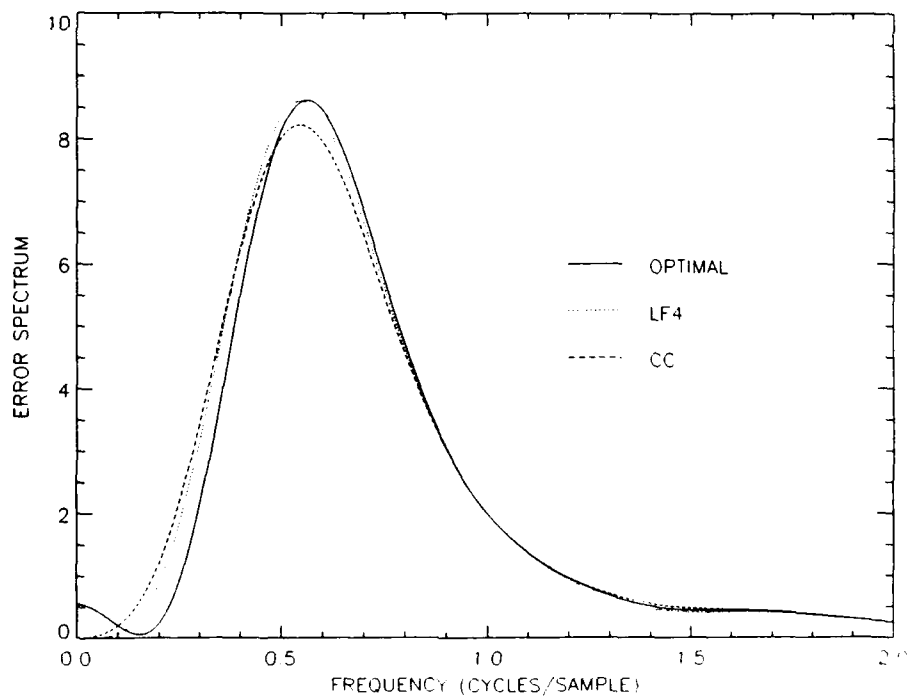
(b)

Fig. 12 — (a) Errors for three four-point interpolators: the $(1/\nu^4)$-optimal, LF4, and CC; (b) error spectrum for three four-point interpolators and a $1/\nu^4$ image spectrum
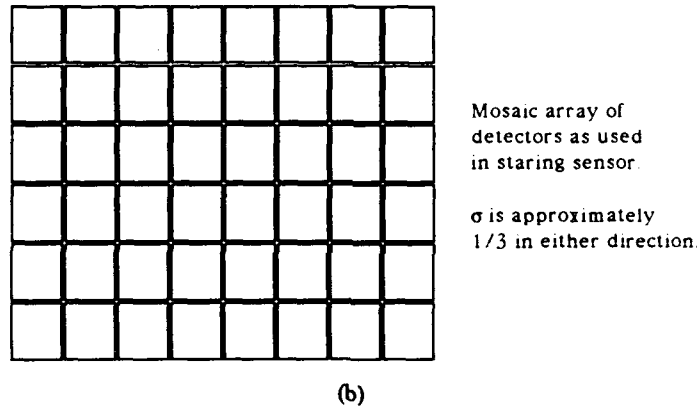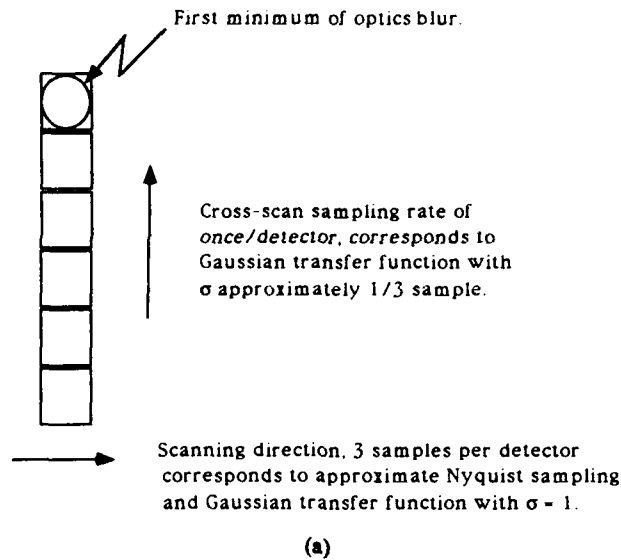
(a)

(b)

Fig. 13 — (a) Linear detector array for a scanning sensor;
(b) two-dimensional array for a staring sensor

The solution of Eq. (55), using Eq. (54) with the spectrum of Eq. (65), is tedious but straightforward. For $N = 2$ the optimal kernel is

$$r(s) = \frac{\gamma^{s^2} - \gamma^{[1+(s-1)^2]}}{1 - \gamma^2} \quad \text{with } \gamma = e^{-(1/2\sigma)^2} \quad (0 \le s \le 1), \tag{67}$$

which is plotted in Fig. 14 for the three values of σ. Like all kernels derived in this report, it is symmetric. Figure 15 plots the analytically complicated $N = 4$ results. Figure 16 shows the image power spectra corresponding to the three values of σ; these spectra have been scaled so that the total energies, i.e., image variances, are identical. Figures 17 and 18 show the $N = 2$ and $N = 4$ error factors, respectively, both for a shift of .25. Notice that for $\sigma = .33$, the performance as N changes from 2 to 4 is small. The square roots of the integrals over the error spectra (Fig. 19) are proportional to the rms errors, and the difference of these proves to be less than 1 percent. On the other hand, if the spectra are truncated at the Nyquist frequency, the error reduction is 13 percent.
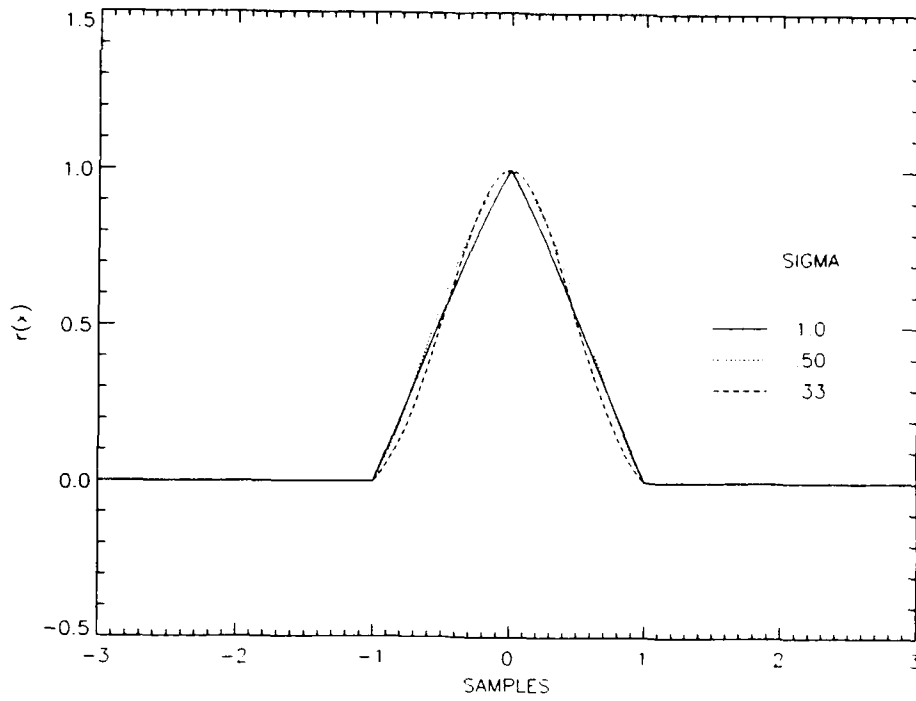
28

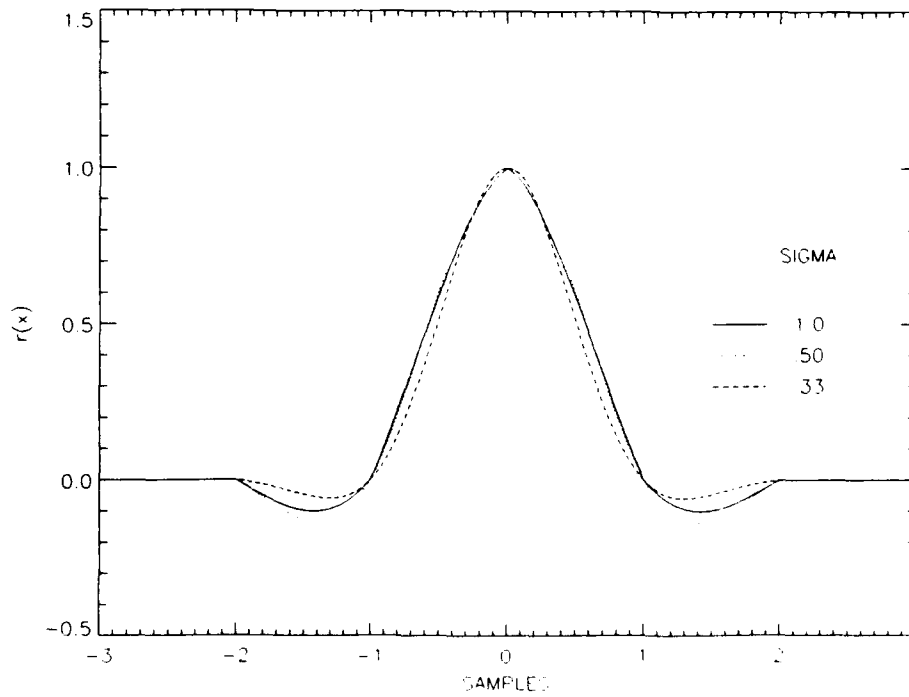Fig. 14 — Two-point Gaussian-optimal kernels
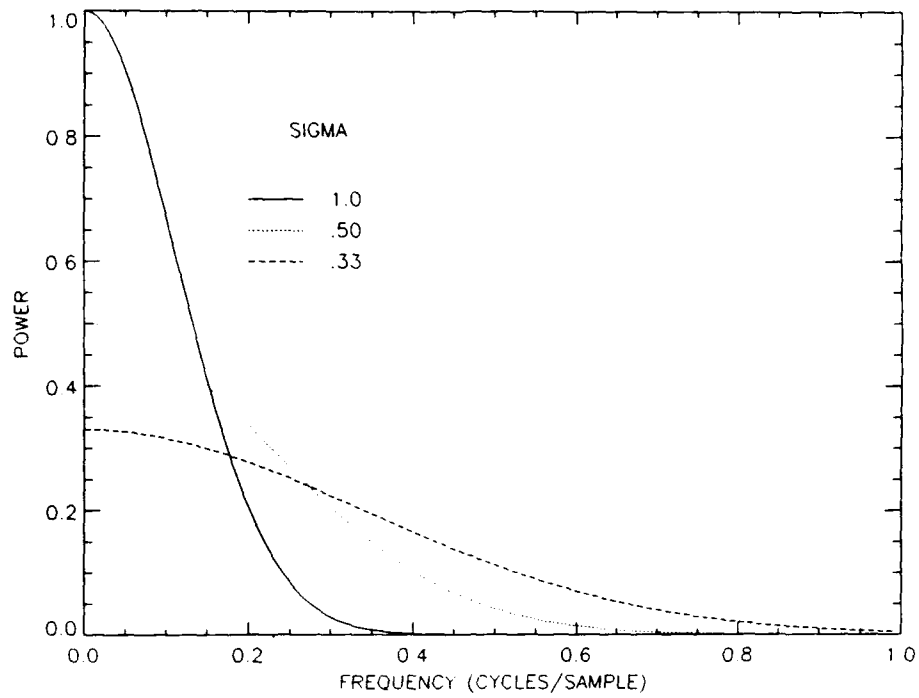

Fig. 15 — Four-point Gaussian-optimal kernels

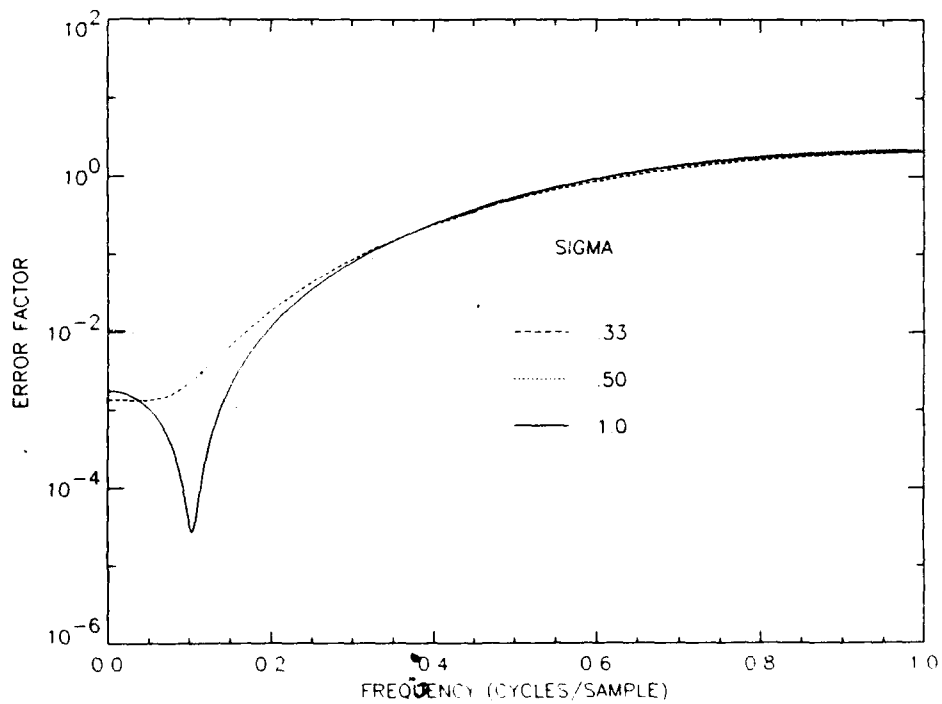Fig. 16 — Gaussian power spectra with identical total energies

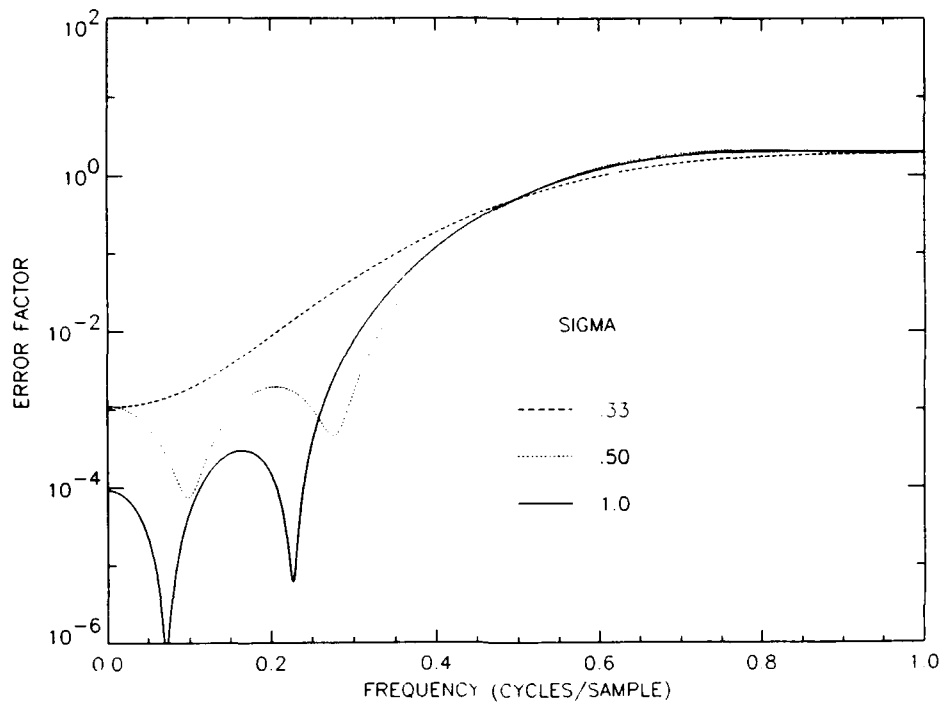Fig. 17 — Error factors for two-point Gaussian-optimal kernels

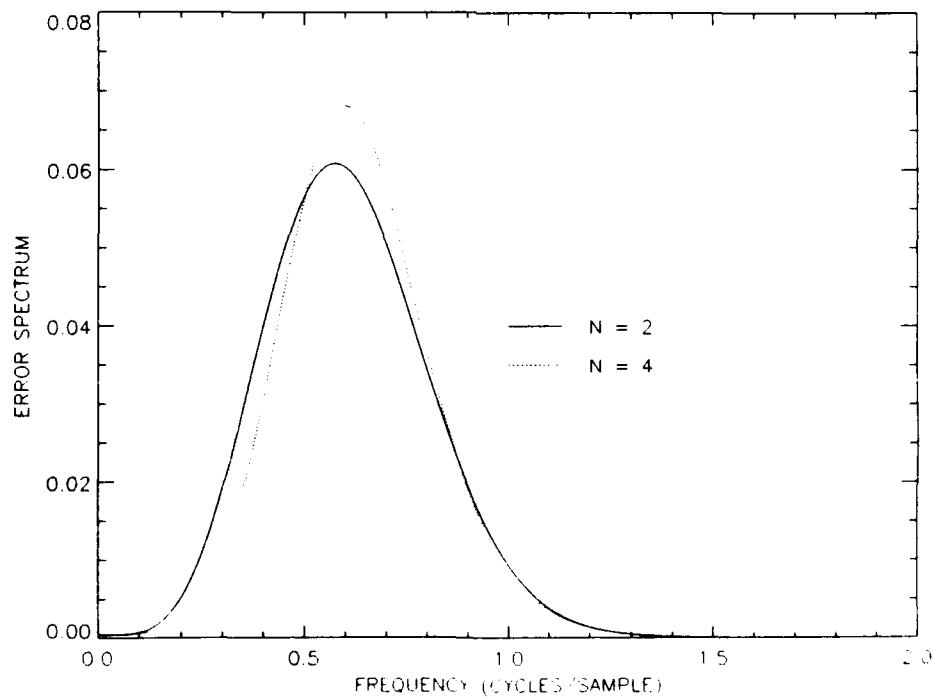Fig. 18 — Error factors for four-point Gaussian-optimal kernels

Fig. 19 — Error spectra for Gaussian-optimal interpolators acting on Gaussian spectrum: σ = .33

Increasing σ, which is measured in samples, can be thought of as either raising the sampling rate or compressing the spectrum toward dc, as shown in Fig. 16, because the unit used here to measure frequency is always cycles/sample. For $\sigma = .5$ samples (Fig. 20), the rms error is reduced by approximately 16 percent as N changes from 2 to 4; for $\sigma = 1$ (Fig. 21), it is smaller by a factor of 3.3. The failure of the four-point method to greatly out perform the two-point for $\sigma = .33$ reflects the difficulty of designing an interpolator with a consistently small error factor over a wide range of frequencies.

Our final comparison is among the various values of σ for a given N. As noted above, the rms error (for $\nu_t = +\infty$) is nearly the same for $N = 2$ and $N = 4$ when $\sigma = .33$. Either case can, therefore, serve as a common baseline. For $N = 2$, the rms error is reduced by a factor of 1.9 as σ changes to .50, and by another factor of 3.7 as σ changes to 1.0. For $N = 4$, the corresponding reductions are 2.58 and 10.3, respectively. Optimal interpolation achieves its greatest gains for narrowband spectra.

Hybrid Designs

Figures 11 and 12(b) demonstrate that polynomial interpolators may be guilty of overkill at low frequencies. Optimal interpolators achieve superior performance by leaving some residual energy near dc. Similar results are evident in the Gaussian error spectra of Figs. 20 and 21. On the other hand, if the image spectrum actually ends short of the Nyquist frequency, then the $\nu_t = +\infty$ optimal kernels can perform poorly because their domain of superior performance is often at high frequencies. Therefore, both polynomial and $\nu_t = +\infty$ optimal solutions have defects when applied to truncated spectra. This problem could be remedied by solving Eqs. (55) and (54) for a spectrum truncated at the Nyquist limit. Such solutions are quite complicated analytically.

An alternative approach uses Eq. (40) to define interpolators that sacrifice some of the low-frequency performance that is present in LF-4, for example, for smaller high-frequency errors. The error factor for LF-4 varies as ν^8 near dc, but image power spectra seldom diverge as fast as $1/\nu^4$. Therefore, the error spectrum is usually of order greater than ν^4, rapidly dying at dc.

By using Eq. (40), two degrees of design freedom in a four-point interpolator can be spent to set the error factor to zero at, say, half the Nyquist frequency, $\nu = 1/4$. The remaining two degrees of freedom can be used to constrain to the value zero: (a) the error factor at dc, and (b) the first derivative of the complex error $E_s(\nu)$ at dc. Two degrees of freedom are required for the first step because two phases of a sinusoid may be present in an image for $\nu = 1/4$. The dc conditions (a) and (b) make $(d/d\nu)^n[e_s^2(\nu)]|_0 = 0$ for $n = 0, ..., 3$, just as for linear interpolation.

Figure 22 shows error spectra out to the Nyquist limit that result from applying three interpolators to a $1/\nu^2$ spectrum. All these may be regarded as four-point interpolators because, as was shown previously, LIN is actually the optimal N-point interpolator for a $1/\nu^p$ spectrum when $p = 2$. Above $\nu = .2$, benefits of the above hybrid design are evident.

Figure 23 shows another set of error spectra, now for a $1/\nu^4$ image spectrum, in which the optimal four-point $p = 2$ interpolator LIN has been replaced by the optimal four-point $p = 4$ interpolator (Eq. (64)). Here the sacrifice in low-frequency performance required to construct the hybrid is severe. The image power is too divergent near dc, and the hybrid solution is inferior.

Figure 24 shows the final hybrid example. The image spectrum is Gaussian with $\sigma = .33$, and the $\nu_t = +\infty$ optimal is compared with LF-4 and the above hybrid. LF-4 overkills low frequencies, the optimal achieves its largest relative gain beyond the Nyquist limit, and the hybrid is the best choice for an image spectrum truncated at the Nyquist.
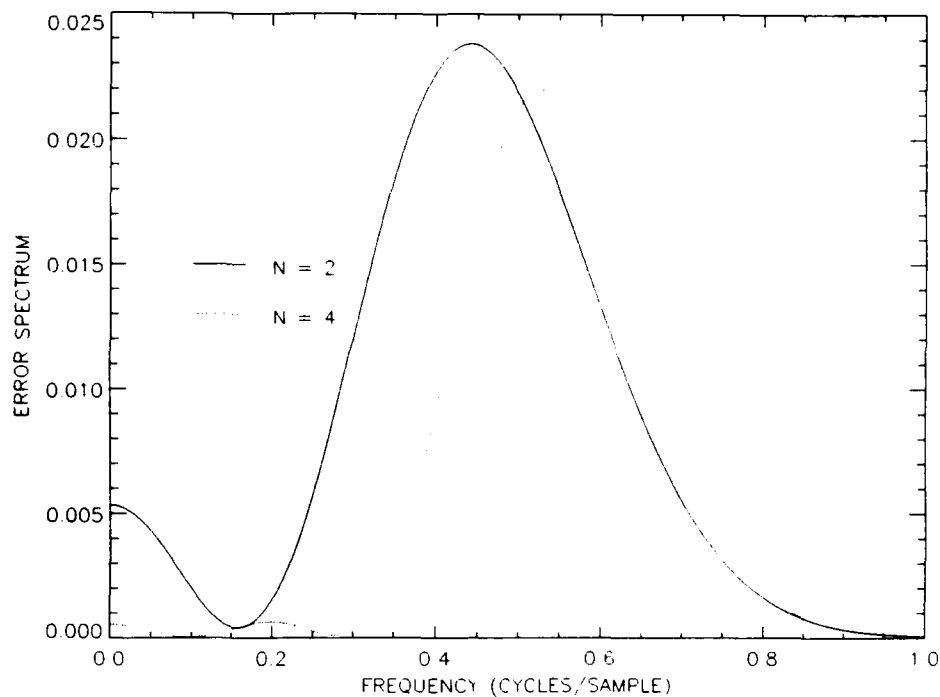
Fig. 20 — Error spectra for Gaussian-optimal interpolators
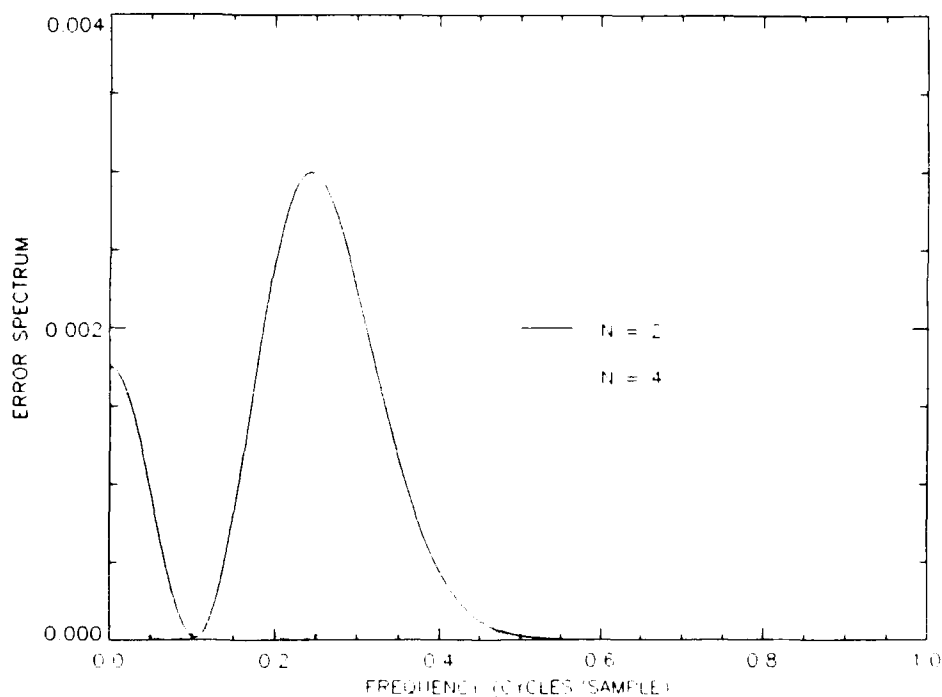acting on Gaussian spectrum: $\sigma = .50$

Fig. 21 — Error spectra for Gaussian-optimal interpolators
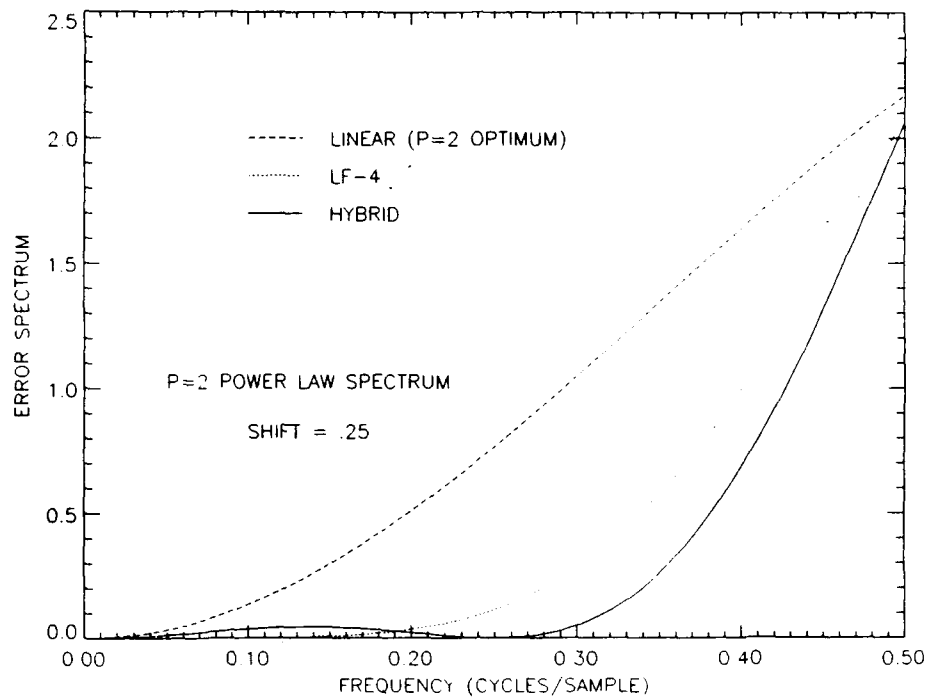acting on Gaussian spectrum: $\sigma = 1.0$

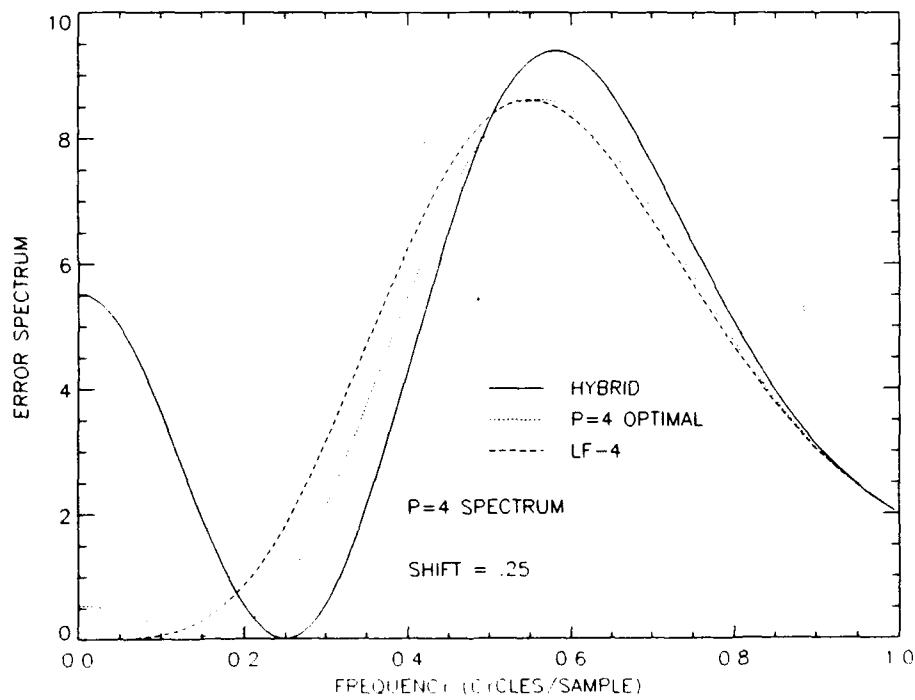Fig. 22 — Error spectra for $N = 4$ interpolators, $p = 2$, shift = .25

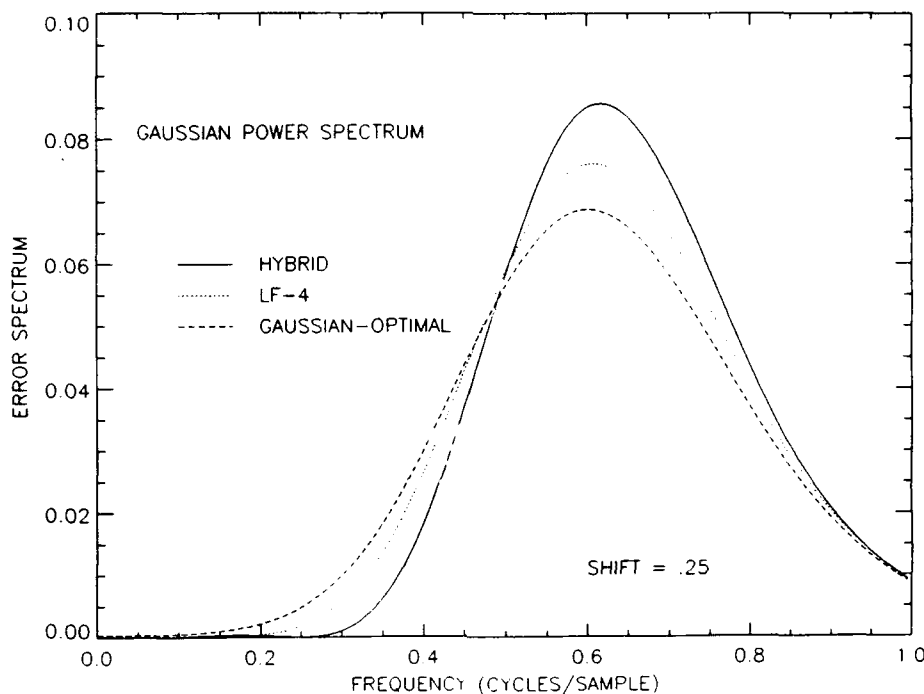Fig. 23 — Error spectra for $N = 4$ interpolators, $p = 4$, shift = .25

Fig. 24 — Error spectra as in Fig. 23, but with $\sigma = .33$ Gaussian input spectrum

Nulls in the error factor or its derivatives could be forced at other frequencies besides $\nu = 1/4$, or even instead of $\nu = 0$. For higher N, larger numbers of such constraints can be imposed, all implemented by setting Eq. (40) or its derivatives to zero at the selected frequencies and solving for the $r(n + s)$. Such choices are dictated in practice by a knowledge of the types of power spectra over which the interpolation methods are to be applied.

8. BEYOND INTERPOLATION

The optimal interpolators described above for orders $N = 2$ and $N = 4$ usually reduce errors by factors of less than two when compared to conventional interpolators of the same order. Larger reductions are achievable by increasing N, and the above methods provide a means for designing such interpolators. For example, the error factor for LF-10 was shown in Fig. 7, and Fig. 25 compares the error spectra after using LF-2,-4,-10 on a $1/\nu^3$ spectrum truncated at $\nu = 1/2$. LF-10 produces a smaller rms error than LIN by a factor of 2.3. Reductions with increasing N generally occur because N equals N_d, the number of design degrees of freedom for the interpolator. However, for certain applications, N_d can be increased more effectively without increasing N.

In change detection, two digital images are often compared by subtraction after one has been overlaid on the other, i.e., resampled. Our analyses are well suited to this technique. The mean square error d_s^2 becomes a measure of residual clutter after subtraction.

The change to be detected may be an extensive one, such as a region of blood flow on an angiogram or large-scale agricultural evolution as seen from space. It may be a local change, such as the motion of a small object. The general goal is to combine the two images so that only such changes persist.

The N-point interpolators described above are actually digital filters operating on an image. To implement change detection with interpolation, a second image is used, but only in the subtraction step.
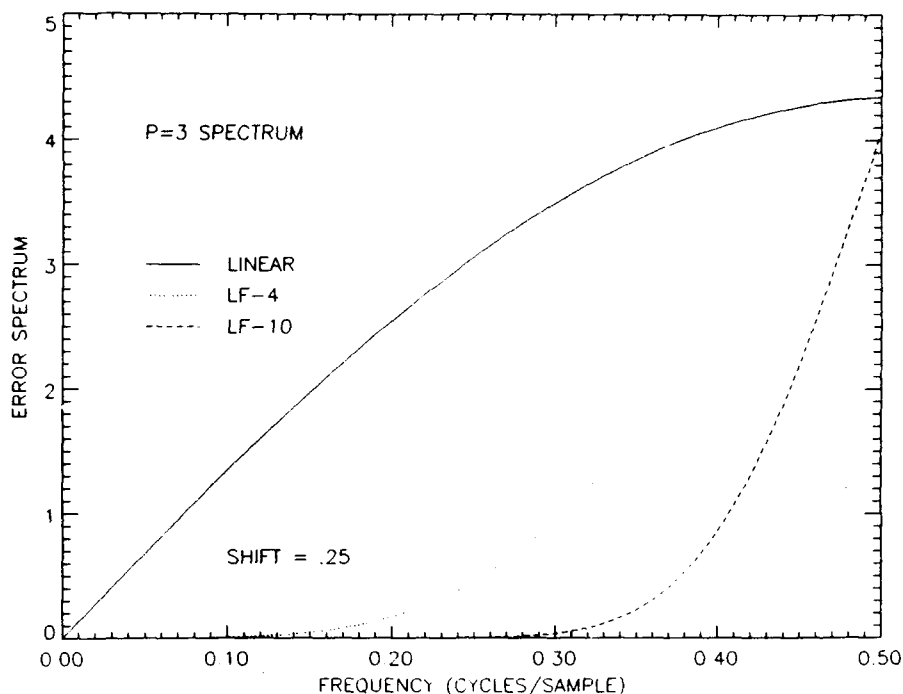
Fig. 25 — Error spectra for LF-2,-4,-10, and p = 3 spectrum, shift = .25

It can be thought of as a template for gauging interpolation accuracy. However, by digitally filtering both images before subtraction, N_d can be increased to nearly $2N$. This dual filtering introduces a new element into the analysis, consisting of constraints on the individual filters to preserve the change that is to be detected. For example, for extensive changes, an appropriate constraint may be to require perfect dc response of each filter. For detection of small moving targets of known shape, the filters can be constrained to conserve target energy or peak amplitude.

Figure 26 illustrates performance results for two examples of this dual difference filtering. These should be compared to Fig. 25 (note the scale change). The dual LF-4 is the analog of the interpolating LF-4. The constraint imposed on its single-image filters was preservation of dc, i.e., of local mean values. The dual hybrid is analogous to the interpolating hybrid described above, with the important new feature that errors at ν = 1/2 are also perfectly suppressed, an impossibility with interpolation. This particular dual hybrid was constrained to preserve the peak amplitude of a small Gaussian target (corresponding to σ = 1 in the Section 7 analysis). The dual LF-4 and hybrid yield smaller rms errors than even LF-10 (Fig. 25) by factors of 5.9 and 12.1, respectively.

Because the second image is now also filtered before subtraction, the difference image is no longer a measure of interpolation error. Dual filtering methods relinquish the goal of interpolation for the sake of maximizing a signal-to-clutter ratio. These ideas are explored further in a companion paper [18].

9. SUMMARY

This report provides tools for evaluating and designing interpolation and resampling methods. It begins by deriving a generalized version of Parseval's Theorem. The theorem has strong and weak forms, depending on whether the image is over- or undersampled. An extension of the derivation produces a second theorem that replaces the sampled image with an error image. Strong and weak versions of this theorem are also proved, but now the strong version applies whenever the *image*—not the error image—is
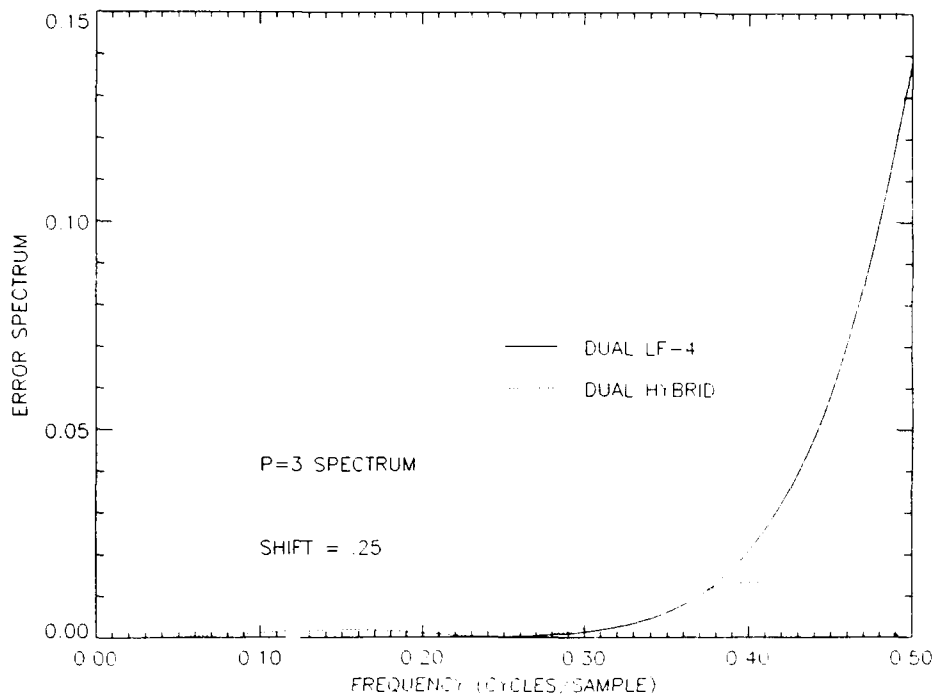
Fig. 26 — Error spectra for dual filtering methods

oversampled. However, even for undersampled imagery, the weak version provides a valuable tool for predicting interpolation errors in an average sense.

The proofs of these theorems lead to the derivation of a fundamental formula for interpolation error, which is expressible as an error spectrum. This can be written as a product of the image power spectrum and an error factor that depends only on the interpolation kernel. Forms for the error factor are derived that are useful for interpolators that are of finite extent either in space or in frequency.

An equation for the minimum-squared-error interpolator is derived and solved for various model image spectra: constant inband, Lorentzian, power law, and Gaussian. When image spectra are modelled with analytic forms extending over infinite frequency range, the optimal solutions often result in residual error spectra concentrated around the Nyquist limit. Such solutions are sometimes inferior to polynomial methods when applied to Nyquist-truncated versions of the model spectra.

Methods are described for customizing the performance of an N-point interpolator when the precise form of the image spectrum is unknown. Performance error can be made zero at selected frequencies, the number of which is limited only by the order of the interpolator. Such hybrid methods can also be designed to keep the error small in selected regions of the spectrum by concentrating many zeros of the error factor into a small range, such as for LF-N, in which all zeros are collapsed to dc. The hybrids are also valuable suboptimal methods when the form of the ideal interpolator is cumbersome, as often happens with image power spectra that are truncated forms of analytic functions.

REFERENCES

1. A. Schaum and M. McHugh, "Analytic Methods of Image Registration: Displacement Estimation and Resampling," NRL Report 9298, February 1991.

2. G.W. Wornell, D.S. Kauffman, and B. Sharpe, "Minimum Mean-Square Error Resampling for Remotely Sensed Imagery," 12th Canadian Symposium on Remote Sensing, *IEEE* **89CH2768-0**, 1989.

3. A.J. Mord, N.H. Endsley, E. Ramberg, and H.J. Reitsema, "Electronic Image Alignment: Implementation and Applications to Imaging System Design," *Opt. Eng.* **24**(3), 507-515 (1985).

4. J. Owczarczyk, W.J. Welsh, and S. Searby, "Performance Analysis of Image Registration Techniques," Third Int. Conf. on Im. Proc. and Applic., 1989, pp. 10-13.

5. J.A. Parker, R.V. Kenyon, and D.E. Troxel, "Comparison of Interpolating Methods for Image Resampling, *IEEE Trans. Med. Im.* **MI-2**(1), 31-39 (1983).

6. A. Ferraioli, "Review of Clinical Applications of Cardiac Digital Angiography," *IEEE Trans. Med. Im.* **9**(1), 113 (1990).

7. K.P. Prasad and P. Satyanrayana, "Fast Interpolation Algorithm using the FFT," *Electron. Lett.* **22**, 185-187 (1986).

8. S. Shlien, "Geometric Correction, Registration, and Resampling of LANDSAT Imagery," *Canadian J. Remote Sensing* **5**, 74-89 (1979).

9. S.K. Park and R.A. Schowengerdt, "Image Sampling, Reconstruction, and the Effect of Sample-scene Phasing," *J. Appl. Opt.* **21**(17), 3142-3151 (1982).

10. S.K. Park and R.A. Schowengerdt, "Image Reconstruction by Parametric Cubic Convolution," *Computer Vision, Graphics, and Image Processing* **23**(N3), 258-272 (1983).

11. R. Lucke, A Local Interpolator Derived from the DFT, submitted for publication in *IEEE Trans. Sig. Proc,* 1991.

12. D. Fraser, "Interpolation by the FFT Revisited — An Experimental Investigation," *IEEE Trans. Acoust. Sp. Sig. Proc.* **37**(5), 665-675 (1989).

13. G. Wolberg, *Digital Image Warping* (IEEE Computer Society Press, Los Alamitos, CA, 1990) pp. 139-40.

14. E.A. Robinson and S. Treitel, *Geophysical Signal Analysis* (Prentice-Hall, Englewood Cliffs, N.J., 1980) p. 163.

15. W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes* (Cambridge Univ. Press, Cambridge, 1986) pp. 47-51.

16. S.M. Kay, *Modern Spectral Estimation: Theory and Application* (Prentice-Hall, Englewood Cliffs, N.J., 1988).

17. M.S. Farber, S.J. Hemple, and B.A. Eckstein, "Characterization of IRAMMP Data Using Scene Registration," Proceedings of the IRIS Specialty Group on Targets, Backgrounds, and Discrimination, Vol. 1, March 1, 1991, ERIM, Sandia National Laboratory, Albuquerque, NM, pp. 89-114.

18. A. Schaum, "Dual Difference Filtering for Change Detection in Digital Signals," submitted for publication in *IEEE Trans. Sig. Proc,* 1992.

Appendix

DERIVATION OF THE COMB TRANSFORM

The fundamental theorem of Fourier series can be written as

$$\hat{f}(v) = \sum_n e^{-2\pi i n v} c_n, \qquad v \in [-1/2, 1/2] \qquad (A1)$$

where

$$c_n = \int_{-1/2}^{1/2} e^{2\pi i n \eta} \hat{f}(\eta) d\eta, \qquad (A2)$$

for any reasonably well-behaved function f [A1]. Therefore, formally defining

$$k(v) = \sum_n e^{-2\pi i n v}, \qquad (A3)$$

Eqs. (A1) and (A2) can be combined into

$$\hat{f}(v) = \int_{-1/2}^{1/2} k(v - \eta)\hat{f}(\eta) d\eta. \qquad (A4)$$

Because Eq. (A4) holds for (rather) arbitrary $\hat{f}$:
  (a) k is by definition the Dirac delta function on the interval $[-1/2, 1/2]$.
However, it is obvious from Eq. (A3) that:
  (b) k is periodic with period one.
Together, (a) and (b) mean that k must be the sum of displaced delta functions. That is:

$$k(v) = \sum_n \delta(v - n). \qquad (A5)$$

The right-hand side of Eq. (A5) is the definition of the comb, so that Eq. (A3) becomes

$$\sum_n e^{-2\pi i n v} = \text{comb}(v). \qquad (A6)$$

However, the left-hand side of Eq. (A6) is just the Fourier transform of the sum of displaced delta functions:

$$\sum_n \delta(x - n) = \text{comb}(x). \qquad (A7)$$

That is, $\text{comb}(v)$ and $\text{comb}(x)$ are Fourier transform pairs, as claimed in Eq. (9). For an alternative proof that treats delta functions as the limit of Gaussians, see Ref. A2.

REFERENCES

A1. W. Rogosinski, *Fourier Series* (Chelsea Pub. Co., New York, 1959).

A2. R.N. Bracewell, *The Fourier Transform and its Applications*, (McGraw-Hill, New York, 1978).